Rialtas na hÉireann
Government of Ireland

Plean Digiteach don Ghaeilge

Teicneolaíochtaí Urlabhra agus Teanga 2023-2027

Arna ullmhú ag an Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meán
www.gov.ie

Clár

Teachtaireacht ón Aire agus ón Aire Stáit ?

Nóta Buíochais ?

Achoimre Fheidhmeach ?

Caibidil 1:
Réamhrá: an Ghaeilge sa Ré Dhigiteach ?

Cuid I **Caibidil 2:**
Anailís Teangeolaíochta agus Doiciméadú ar an Ghaeilge Labhartha agus Scríofa ?

Caibidil 3:
Corpais ?

Caibidil 4:
Bunachair Faisnéise Teangeolaíochta ?

Caibidil 5:
Uirlisí Próiseála Teanga Nádúrtha ?

Caibidil 6:
Giniúint Teanga Nádúrtha ?

Caibidil 7:
Samhlacha Urlabhra don Ghaeilge ?

Cuid II **Caibidil 8:**
Sintéis Chainte ?

Caibidil 9:
Aithint Cainte ?

Caibidil 10:
Córais Chomhrá Ríomhbhunaithe ?

Caibidil 11:
Meaisínaistriúchán ?

Caibidil 12:
Aisghabháil Faisnéise ?

Cuid III **Caibidil 13:**
Áiseanna don Phobal Mór ?

Caibidil 14:
Feidhmeanna Oideachasúla ?

Caibidil 15:
Áiseanna: Míchumas agus Rochtain ?

Caibidil 16:
Conclúid: an Plean Digiteach a Chur i bhFeidhm ?

Aguisín A ?

Teachtaireacht ón Aire agus ón Aire Stáit

Mairimid inniu i 'Ré Dhigiteach' ina n-úsáidtear an teicneolaíocht go forleathan i ngnéithe éagsúla den saol. Tá Éire ar cheann de lárionaid mhóra na nuatheicneolaíochta san Eoraip le tamall de bhlianta anuas, ach is é an Béarla atá in uachtar sa réimse sin trí chéile agus an Ghaeilge i riocht i bhfad níos laige. Tá ár dteanga náisiúnta ar cheann de na seoda is luachmhaire atá ag cultúr agus oidhreacht na hÉireann agus is léiriú í ar an bhféiniúlacht ar leith atá againn mar náisiún. Is mithid dúinn díriú anois ar an oidhreacht sin a chónascadh le sochaí dhigiteach na linne seo.

Treochlár atá sa Phlean seo don fhís atá le réadú sna blianta seo romhainn. Is é an cuspóir atá leagtha romhainn againn go gcuirfear ar chumas phobal na Gaeilge an teanga a chothú agus a athbheochan trí earnáil bhríomhar teicneolaíochta Gaeilge a bhunú agus a fhorbairt, ionas go mbeidh deiseanna úra ag an bpobal sin a bheith rannpháirteach go hiomlán i sochaí nua-aimseartha na linne seo. Forbrófar teicneolaíochtaí sa réimse urlabhra agus teanga go háirithe, ionas go mbeidh ár dteanga gob ar ghob le mórtheangacha na hEorpa ó thaobh feidhm agus úsáid na ngnéithe sin den nuatheicneolaíocht. Má éiríonn linn sa mhéid sin, cuideofar go mór le húsáid na Gaeilge a normalú sa tsochaí trí chéile. Go deimhin, is í an tsamhail atá againn go dtiocfaidh an teanga faoi bhláth san aois dhigiteach má thapaítear na deiseanna a chuireann an nuateicneolaíocht ar fáil.

Is plean náisiúnta é an Plean Digiteach a thagann le tosaíochtaí an Aontais Eorpaigh ó thaobh an aistrithe dhigitigh de. Tagann an Plean seo freisin leis an Straitéis 20 Bliain don Ghaeilge 2010-2030, go háirithe Réimse 5 den Straitéis: Na Meáin agus an Teicneolaíocht. Ullmhaíodh an cháipéis seo freisin i gcomhthéacs beartais ghaolmhara agus reachtaíocht ghaolmhar eile, lena n-áirítear IS - Is Cuid den Saol é Anois: Straitéis Náisiúnta um Intleacht Shaorga d'Éirinn; Leas a bhaint as Digitiú: Creat Digiteach na hÉireann; Plean Gníomhaíochta don Ghaeilge, 2018-2022; agus an tAcht Míchumais, 2005.

Tugtar léargas sa Phlean Digiteach seo ar an taighde atá de dhíth chun go mbeidh acmhainní teangeolaíochta Gaeilge ar an bhfód sna blianta seo romhainn. Léirítear freisin na bunteicneolaíochtaí urlabhra agus teanga a bhaineann leis na hacmhainní teangeolaíochta sin agus déantar cur síos ar na háiseanna agus na feidhmchláir is gá a fhorbairt amach anseo. Díreofar go háirithe ar dhaoine atá faoi mhíchumas, grúpa leochaileach a fhágtar ar lár, uaireanta. Tá sé leagtha romhainn againn go mbeadh rochtain ionchuimsitheach ag aon duine ar spéis leo an teanga ar thacaíocht agus ar áiseanna ionas gur féidir leo an Ghaeilge a úsáid go furasta sa réimse digiteach ar bhonn laethúil. Dá réir sin, tacaíonn an Plean le teicneolaíochtaí agus le háiseanna úrscothacha i réimsí éagsúla: teicneolaíocht urlabhra ilchanúnach, innill chuardaigh fheabhsaithe, meaisínaistriúchán, ardáin agus aipeanna oideachasúla a chabhróidh le teagasc na teanga mar aon le léitheoirí scáileáin, sintéis chainte agus áiseanna eile do lucht labhartha na Gaeilge a bhfuil lagú radhairc orthu, gan ach roinnt díobh a lua.

Beidh tacaíocht thrasrannach agus iarracht idirdhisciplíneach chomhthreomhar chomhtháite ag teastáil chun an Plean seo a chur i ngníomh. Déanfaidh teangeolaithe, innealtóirí teanga agus saineolaithe eile a gcion féin leis an obair fhiúntach atá ar na bacáin a chur i gcrích. Beidh oifigigh na Roinne i mbun oibre chun na féidearthachtaí uile a bhaineann le feidhmiú an Phlean a chíoradh, ina measc, na deiseanna le struchtúir nua a chur ar bun, trína mbeidh na teicneolaíochtaí atá á bhforbairt sna hinstitiúidí éagsúla á maoirsiú agus á dtreorú ar bhonn comhtháite. De réir mar a thiocfaidh na teicneolaíochtaí nua in inmhe i gcomhthéacs na Gaeilge, táimid ag súil go mbainfear úsáid astu i ngnólachtaí na hearnála príobháidí ar mhaithe le seirbhís i nGaeilge a chur ar fáil do na húsáideoirí deiridh. Beidh ról gníomhach ag pobal na Gaeilge i ngach céim de chur i bhfeidhm an Phlean ionas go mbeidh an taighde á threorú ag riachtanais na n-úsáideoirí.

Is mionteanga uathúil í an Ghaeilge toisc an chosaint bhunreachtúil atá aici ón Stát mar an teanga náisiúnta agus toisc í a bheith aitheanta mar theanga oifigiúil de chuid an Aontais Eorpaigh. Dá chionn sin, measaim go mbeidh an Plean seo ina eiseamláir d'fhorbairt agus neartú mionteangacha eile amach anseo.

Tuigimid ar fad go dtagann athrú agus forás go síoraí seasta ar an teicneolaíocht. De réir mar a fhorbrófar an taighde agus mar a eascróidh torthaí agus áiseanna as an taighde sin, beidh an Plean le hathbhreithniú go foirmiúil agus le tabhairt cothrom le dáta uair gach cúig bliana. Anuas ar na hathbhreithnithe foirmiúla, coinneofar súil ghrinn i gcaitheamh an ama ar a bhfuil ag titim amach sa réimse seo, agus beifear in ann na tosaíochtaí a leasú de réir mar is gá.

Táimid ag tnúth go mór le bheith ag obair i gcomhar lenár bpáirtithe leasmhara agus le pobal na Gaeilge trí chéile sna blianta seo romhainn chun an fhís agus an sprioc uaillmhianach atá leagtha amach againn sa Phlean seo a thabhairt i gcrích ar mhaithe le fás agus forbairt na teanga sa réimse ríthábhachtach seo dár saol.

Catherine Martin TD
An tAire Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meán

Jack Chambers TD
Príomh-Aoire an Rialtais agus an tAire Stáit Gaeltachta agus Spóirt

Nollaig 2022

Nóta Buíochais

Forbraíodh an Plean seo de thoradh obair den scoth a rinne taighdeoirí cumasacha a choimisiúnaigh an Roinn. Táimid faoi chomaoin ag na húdair: Ailbhe Ní Chasaide, Neasa Ní Chiaráin, Elaine Uí Dhonnchadha i gColáiste na Tríonóide, Baile Átha Cliath; Teresa Lynn agus John Judge in Ollscoil Chathair Bhaile Átha Cliath.

Le linn na hoibre seo, fuair na húdair comhairle ar ghnéithe den teicneolaíocht agus aiseolas ar dhréachtaí den Phlean ó shaineolaithe eile in Éirinn agus i gcéin. Táimid fíorbhuíoch díobh ar fad, agus go háirithe: Brian Ó Raghallaigh, Mícheál Ó Meachair, Brian Davis, Monica Ward in Ollscoil Chathair Bhaile Átha Cliath; Emily Barnes, Andy Murphy, Harald Berthelsen, Liam Lonergan, Christer Gobl i gColáiste na Tríonóide, Baile Átha Cliath; Kevin Scannell in Ollscoil St. Louis i Meiriceá; Mengjie Qian in Ollscoil Cambridge; Peter Bell, Simon King, Steve Renals agus Christian Veaux san Ionad Taighde ar Theicneolaíocht Urlabhra, Ollscoil Dhún Éideann; Martin Russell in Ollscoil Birmingham; Shane Ó hEidhin, Micheál Ó Maolruanaidh i bhForas na Gaeilge.

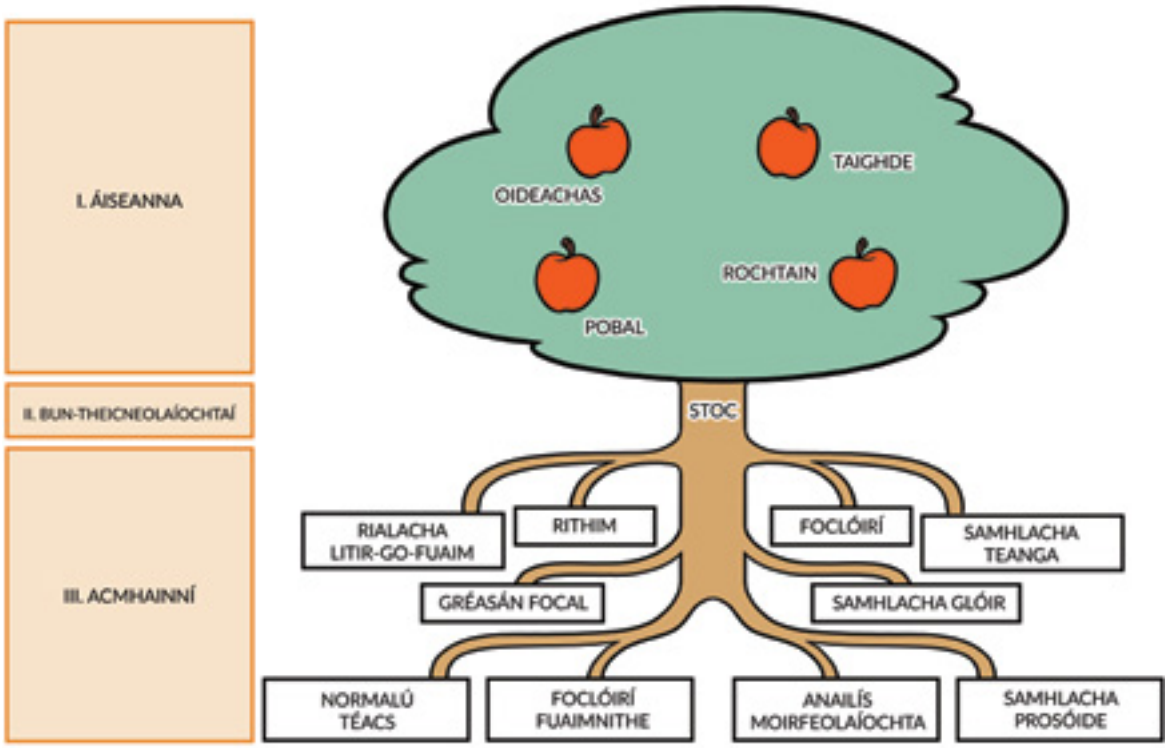
Achoimre Fheidhmeach

Tá an teicneolaíocht urlabhra agus teanga i gceartlár ár saoil – ár saol oibre, ár saol sóisialta, an chaoi a dtugaimid faoin bhfoghlaim, an dóigh a dtagaimid ar fhaisnéis, ár gcaitheamh aimsire, agus mar sin de. Tá an Béarla ceannasach sa réabhlóid dhigiteach agus deirtear gur 'buama ama digiteach' atá sa réabhlóid sin do na teangacha neamhfhorleathana, agus dlús á chur leis an ráta tubaisteach ag a gcaillltear teangacha. Braitheann an seachadadh teanga ar an nglúin óg, a bhfuil taithí mhaith acu ar an teicneolaíocht; gan amhras ar bith, beidh soláthar teicneolaíochta dár dteanga féin ríthábhachtach má tá an Ghaeilge le teacht slán. Ar an taobh dearfach den scéal, le forbairt na teicneolaíochta urlabhra agus teanga, agus an taighde teangeolaíochta mar bhonn taca léi, tá deiseanna úra den scoth ann teangacha a chothú agus a athbheochan. Dá bhrí sin, is snáithe tábhachtach amháin é an Plean Digiteach i gcur chuige an Rialtais don Ghaeilge.

Leagtar amach sa cháipéis seo an taighde agus an fhorbairt a bheidh ag teastáil chun an Ghaeilge a thabhairt isteach sa ré dhigiteach. Beidh gá le hiarracht idirdhisciplíneach a mbeidh teangeolaithe, innealtóirí, teangeolaithe ríomhaireachta, eolaithe urlabhra, agus speisialtóirí bogearraí páirteach ann. Is den tábhacht é go ndéanfar an obair sin i ndlúthchomhar le pobal na Gaeilge, agus le páirtithe leasmhara i réimsí amhail oideachas, míchumas agus rochtain.

Tá na réimsí sonracha feidhme is tábhachtaí don Ghaeilge éagsúil leis na réimsí a dtugtar tús áite dóibh i dteangacha ar nós an Bhéarla, toisc gur ar chúinsí tráchtála den chuid is mó a bhíonn an fhorbairt bunaithe i gcás na mórtheangacha. Ní mór na tosaíochtaí i gcás na Gaeilge a dhíriú ar na riachtanais is práinní atá ag pobal na Gaeilge agus ar na spriocanna a ghabhann le slánú na teanga. Dá bhrí sin, is treoirphrionsabal den Phlean Digiteach é gur chóir na dúshláin agus na réitigh a bhfuiltear ag tabhairt fúthu a chur in oiriúint do chomhthéacs na Gaeilge agus na hÉireann agus gur chóir iad a bhunú ar eolas domhain ar an teanga.

Trí cinn de shnáitheanna gníomhaíochta atá i gceist leis an taighde agus leis an bhforbairt atá beartaithe sa Phlean Digiteach; cuirtear an cur chuige sin i láthair



Fíor 1: Na 3 shnáithe (Cuid I, Cuid II agus Cuid III) den taighde agus den fhorbairt a leagtar amach sa Phlean Digiteach

sna trí chuid den doiciméad seo, agus tá samhail den chur chuige sin le feiceáil i bhfoirm crainn i bhFíor 1. Is iad sin I Acmhainní, II Bunteicneolaíochtaí agus III Feidhmeanna. Clúdaíonn na trí shnáithe sin an raon iomlán taighde idir bhuntaighde agus taighde feidhmeach agus ní mór na trí cinn acu a chur ar aghaidh i gcomhthreo lena chéile.

I gcás a lán de léitheoirí an Phlean seo, is iad Cuid II agus Cuid III, ina bpléitear na bunteicneolaíochtaí agus na háiseanna/feidhmeanna sa saol mór, na codanna is soléite, agus is túsphointí maithe iad don ábhar spéise seo. I gCuid I tugtar faisnéis theicniúil mhionsonraithe maidir leis na hacmhainní bunúsacha ar a bhfuil an Plean bunaithe. Is díol spéise an chuid seo dóibh siúd atá ag lorg tuiscint níos iomláine ar an réimse, agus tá soiléiriú le fáil inti ar ghnéithe éagsúla de na forbairtí teicneolaíochta a phléitear i gCuid II agus i gCuid III ina dhiaidh sin. Tugtar cuid mhór faisnéis mhíniúcháin ar na teicneolaíochtaí agus ar a n-ábharthacht le súil is go mbeidh sé ina bhonn eolais agus ina spreagadh dóibh siúd ar mian leo tacú leis an saothar seo agus a bheith páirteach ann.

Sna trí chuid den Phlean, déantar na teicneolaíochtaí agus na hacmhainní sonracha a mhíniú agus a phlé sna caibidlí aonair. Is é atá i gceist anseo an pobal a chur ar an eolas agus spéis na léitheoirí a mhúscailt, go háirithe iad siúd ar spéis leo, b'fhéidir, a bheith páirteach sa réimse nua fáis seo. Is doiciméad fada go leor atá anseo, dá bhrí sin. Mar sin féin, tá sé leagtha amach ionas go dtig leis an léitheoir léargas achomair a fháil ar an réimse seo trí chéile agus breathnú go mion ansin ar ghnéithe sonracha inspéise. Ar an gcúis sin, cuireadh an fhormáid chéanna ar na caibidlí nuair ab fhéidir sin a dhéanamh; cumhdaítear na ceisteanna seo a leanas i gcás gach aon topaic, ag bogadh ón bplé simplí ginearálta ar dtús go plé níos mionsonraithe.

1. Cad é?
2. Cén tábhacht atá leis agus cé dó é?
3. Conas a oibríonn sé?
4. A bhfuil déanta i leith na Gaeilge go dtí seo
5. Obair amach anseo
6. Moltaí

Ag deireadh gach caibidle tugtar leabharliosta ina gcuimsítear na tagairtí a luaitear sa chaibidil sin.

Sa Réamhrá agus sa Chonclúid (Caibidlí 1 agus 16), féachtar ar an mórphictiúr, agus ar na gníomhartha a chaithfear a chur i gcrích ionas go mbeidh an tionchar is mó is féidir ag an bPlean Digiteach ar an teanga agus ar phobal na teanga. Mar a luadh thuas, ní mór, san fhorbairt a bheidh le déanamh, comhthéacs na Gaeilge a chur san áireamh, agus ní mór tús áite thabhairt do na háiseanna siúd atá de dhíth ar phobal na teanga, do na háiseanna lena bhféadfaí feabhas as cuimse a chur ar theagasc agus foghlaim na Gaeilge, agus do na háiseanna lena ndéanfar cách a ionchuimsiú.

Níor cheart smaoineamh ar an bPlean mar thasc teoranta nach bhfuil i gceist ann ach ítimí áirithe teicneolaíochta a sholáthar, ach mar infheistíocht a leanfar di i bhfad na haimsire agus ina gcuimseofar:

- Daoine: saineolaithe sa réimse: tosaíocht is ea oideachas a chur ar thaighdeoirí oilte – dá chionn sin beifear in ann na teicneolaíochtaí agus na bunchlocha teanga a ghabhann leo a fhorbairt ar bhonn leanúnach amach anseo. Is taighdeoirí iad seo a mbeidh ardchumas Gaeilge iontu agus a bheidh in ann oibriú sa timpeallacht idirdhisciplíneach atá luaite thuas.
- Ionaid Bairr Feabhais: ba chóir go mbeadh bonneagar fisiciúil agus teicniúil sna hionaid seo agus foirne idirdhisciplíneacha atá dírithe ar an nGaeilge ag obair iontu. Le saothar na bhfoirne seo tiocfaidh an forás teangeolaíochta agus teicneolaíochta as a n-eascróidh teicneolaíocht urlabhra agus teanga úrscothach don Ghaeilge.
- Moil Nuálaíochta Digití: de réir mar a thiocfaidh an teicneolaíocht agus na feidhmchláir chun cinn, beidh breis deiseanna fós ann le haghaidh na nuálaíochta agus le haghaidh fiontair thráchtála. Tá pleanáil de dhíth anseo, mar shampla gnólachtaí campais a bhunú agus iad a bhraisliú le chéile i Moil Dhigiteacha. Ba dheacair an Ghaeltacht a shárú mar ionad faoi choinne tionscnaimh dá leithéid, ach ba chóir fiontair in áiteanna eile a spreagadh chomh maith. Beidh dlúthnasc idir na gnólachtaí/na moil agus na hionaid taighde tábhachtach ionas go mbeidh na gnólachtaí/na moil gob ar ghob leis na forbairtí is déanaí sa réimse seo, a bhíonn ag athrú de shíor. Ba cheart treoirlínte beartais a ullmhú, agus moltar na brabúis ón taighde agus ón nuálaíocht a eascróidh as an bPlean Digiteach a athúsáid chun tacú le taighde agus le forbairt fhadtréimhseach sa réimse seo.
- Rannpháirtíocht an Phobail: an rath a bheidh ar an bPlean Digiteach, beidh sé ag brath ar a mhéid a bheidh pobal na Gaeilge rannpháirteach ann – sa Ghaeltacht, ar fud na tíre, i measc foghlaimeoirí agus cainteoirí líofa araon, lena n-áirítear iad siúd a fágadh ar lár roimhe seo toisc iad a bheith faoi mhíchumas, agus na daoine ar fud an domhain a bhíonn ag plé leis an teanga ar bhealach ar bith. De thairbhe na teicneolaíochta urlabhra agus

teanga beifear in ann nasc a chruthú idir na pobail éagsúla ar cuid de mhórphobal na Gaeilge iad, ionas go mbainfidh siad, le teann comhfhuinnimh, lántairbhe as féidearthachtaí uile na ré digití don Ghaeilge. Chuige sin, ní mór dúinn dul níos faide ná na gnáthghníomhaíochtaí for-rochtana taighde: ní mór dúinn rannpháirtíocht an phobail san iarracht seo a chothú ón gcosmhuintir aníos, agus iad ina gcomhpháirtithe gníomhacha sna gnéithe éagsúla den taighde agus den fhorbairt agus i scaipeadh na n-aschur.

- Rannpháirtíocht na Gaeltachta: Tá an pobal cainteoirí dúchais sa Ghaeltacht ríthábhachtach ó thaobh chur i gcrích an Phlean Dhigitigh agus táthar ag súil go mbeidh an pobal sin i gceartlár an taighde ildisciplínigh atá le tarlú amach anseo. I measc na meicníochtaí a mholtar chun rannpháirtíocht an aosa óig a chinntiú tá gníomhaíochtaí oideachais agus tionscnaimh thráchtála a lonnú ar na suíomhanna céanna.
- Athbhreithniú rialta: tá an teicneolaíocht ag dul chun cinn go tapa agus, dá bhrí sin, beidh gá le hathbhreithniú leanúnach, agus na spriocanna a leasú de réir mar is gá, chun éabhlóid na teicneolaíochta agus an staid reatha iarbhír a chur san áireamh. Moltar athbhreithniú foirmiúil a dhéanamh ar an bPlean Digiteach uair gach cúig bliana.
- Cistiú: Chun spriocanna an Phlean Dhigitigh a bhaint amach, beidh gá le hinfheistíocht mhórscála, leanúnach agus fhadtéarmach, mar aon le tiomantas sna Ranna Rialtais, ionas go mbeidh na haschuir ón taighde agus ón bhforbairt ar fáil do gach earnáil de phobal na Gaeilge, agus tacaíocht leanúnach á tabhairt dóibh. San iarracht sin, táthar ag súil le tacaíocht agus cistiú airgid ó na gnólachtaí móra atá ag obair san earnáil seo in Éirinn, mar chuid den fhreagracht shóisialta chorparáideach atá orthu – tacaíocht a mbeidh de thoradh uirthi go bhforbrófar earnáil dhúchasach áitiúil teicneolaíochta urlabhra agus teanga, í fréamhaithe go daingean i bpobal na Gaeilge agus ag tarraingt ar thuiscint dhomhain ar an teanga.

I gcomhthéacs na dteangacha ar domhan atá i mbaol, bhí an t-ádh ar an nGaeilge go dtí seo ó thaobh an aitheantais agus na tacaíochta a thug an Stát di. Tá rún againn an Plean Digiteach don Ghaeilge a bheith ina eiseamláir do theangacha mionlaigh agus neamhfhorleathana eile, agus de réir mar a fhorbrófar an Plean, déanfar ár gcuid eolais, ár gcuid taithí agus, nuair is cuí, ár gcuid acmhainní a chomhroinnt le pobail eile atá ag streachailt lena dteanga féin a thabhairt slán i dtimpeallacht luaineach na ré digití.

Ach é a chistiú i gceart, é a bhainistiú go ciallmhar, agus an fhís fhadtéarmach a ghabhann leis a chur romhainn mar sprioc, d'fhéadfadh an Plean Digiteach pobal na Gaeilge a chumasú agus a mhúscailt i dtreo athbheochan, chaomhnú agus dhoiciméadú na teanga. Is í ár dteanga an tseoid chultúrtha is luachmhaire dá bhfuil againn. Tá ceann de na traidisiúin liteartha is sine san Eoraip agus ár n-oidhreacht amhránaíochta, filíochta agus scéalaíochta fite fuaite sa teanga. Agus an teanga á tabhairt isteach sa ré dhigiteach, tá sé mar aidhm ag an bPlean seo ár n-oidhreacht a lánpháirtiú i saol na linne seo in Éirinn, agus ár bhféiniúlacht a choimeád slán dá réir i gcultúr domhanda atá ag teacht faoi bhagairt ón aonchineálacht.

Caibidil 1
Réamhrá: An Ghaeilge sa Ré Dhigiteach

1.1 Cúlra agus cuspóirí

Tá an teicneolaíocht dhigiteach cumarsáide lárnach i saol an lae inniu, agus baineann muid úsáid as san obair, inár saol sóisialta, nuair a bhíonn muid ag foghlaim, ag cuardach eolais, mar chaitheamh aimsire agus san idirghabháil eadrainn agus an saol mór. Ceadaíonn teicneolaíochtaí na hurlabhra agus teanga dúinn déileáil le ríomhairí agus áiseanna digiteacha trí théacs, agus go minic anois, trí mheán na cainte.

Tá an réabhlóid dhigiteach mar a bheadh "digital timebomb" ann do na mionteangacha [1], trína bhfuil cultúr an Bhéarla á scaipeadh ar fud na cruinne – rud a chuireann leis an ráta fíochmhar lena bhfuil mionteangacha á gcailliúint sa lá atá inniu ann. Ag an am céanna, tugann na teicneolaíochtaí digiteacha go leor bealaí dúinn leis an dainséar seo a iompú ina leas – bealaí le doiciméadú, caomhnú agus athbheochan a dhéanamh ar theangacha agus ar chanúintí atá i mbaol (féach mar shampla [2]). Mar go bhfuil an teicneolaíocht fite fuaite le gach gné dár saol, caithfidh sé a bheith ar fáil sa Ghaeilge chun gur féidir an teanga a úsáid sna gnéithe saoil seo uile, agus chun go mbeidh buntáistí na ré digití ar fáil don teanga.

Sa Phlean Digiteach, cuirtear fís i láthair faoin dóigh ina bhféadfaí teicneolaíocht digiteach urlabhra agus teanga a fhorbairt don Ghaeilge. Mínítear an taighde agus an fhorbairt atá de dhíth, agus tugtar léirmheas ar an mhéid atá déanta go dtí seo. Tá an bhéim ar an teanga bheo agus ar na teicneolaíochtaí a chuideoidh le daoine an teanga a fhoghlaim agus a úsáid i ngach réimse dá saol. Tá sé mar bhunphrionsabal don Phlean, go cuirfear na teicneolaíochtaí agus na háiseanna bunaithe orthu ar fáil go forleathan do phobal na Gaeilge.

Beifear ag díriú ar theicneolaíochtaí agus ar acmhainní ar ardchaighdeán, atá ar chomhchéim leis an mheid atá ar fáil sna mórtheangacha. Cuireann an Plean mar sin leis an Straitéis Náisiúnta um Intleacht Shaorga d'Éirinn 2021 "IS - Is Cuid den Saol é Anois", agus léiríonn sé chomh riachtanach is atá sé an teicneolaíocht a bheith ar fáil don Ghaeilge, seachas a bheith bunaithe ar an Bhéarla amháin. Tá an teicneolaíocht ag forbairt go tapa , agus is mithid dúinn a bheith ag smaoineamh ar an taighde fhadtéarmach a bheidh ag teastáil. Fágann sin go bhfuil na haidhmeanna atá againn inniu le huasdátú go rialta agus moltar go ndéanfaí athbhreithniú ar an phlean gach cúig bliana.

Tá léiriú ginearálta sa cháipéis seo ar na réimsí ina bhfuil taighde agus forbairt de dhíth. Tá sé ríthábhachtach go ndéantar pleanáil don teicneolaíocht ar bhealach a fhreagraíonn do chomhthéacs na Gaeilge sa lá atá inniu ann, agus go dtuigtear cé mar is féidir leis an teicneolaíocht tacú leis an teanga agus le pobal na Gaeilge. Dá bhrí sin, tugtar go leor eolais ar theicneolaíochtaí éagsúla agus ar a dtábhacht i gcomhthéacs na Gaeilge, le súil go spreagfaidh sé léitheoirí chun tacú leis an tsaothar agus páirt a ghlacadh san fheachtas. Mar thoradh air seo, is cáipéis sách fada í seo, ach cé go bhfuil plé teicniúil i gceist in amanna, tá sí scríofa ar an iomlán le bheith intuigthe ag an ghnáthléitheoir gan cúlra sa réimse. Déantar iarracht tuiscint bhunúsach a thabhairt ar an ábhar, agus aidhmeanna an Phlean agus na dúshláin romhainn a mhíniú.

Is fiú a lua nach ionann i gcónaí na háiseanna a bheidh tábhachtach don Ghaeilge agus na cinn is tábhachtaí do na mórtheangachaí mar an Bhéarla, ina bhfuil an bhéim ar ghnó agus ar bhrabús. Beidh tosaíochtaí don Ghaeilge dírithe ar na teicneolaíochtaí is mó dá bhfuil gá ag pobal na Gaeilge agus ar na háiseanna is mó a thacaíonn le caomhnú na teanga.

Tá an cuntas ar an taighde agus ar an fhorbairt atá le déanamh leagtha amach i dtrí pháirt. Clúdaítear i gCuid I an taighde ar acmhainní teangeolaíochta srl. atá de dhíth. Baineann Cuid II leis na bun-teicneolaíochtaí urlabhra agus teanga a mbaintear úsáid astu sna hacmhainní teangeolaíochta. I gCuid III, tá cur síos ar na háiseanna agus na feidhmchláir is féidir a thógáil – áiseanna a bhaineann úsáid as aschuir Chodanna I

agus II. Is trí na háiseanna go mórmhór a théann an teicneolaíocht i bhfeidhm ar an phobal.
Sa Réamhrá seo agus sa Chonclúid (Caibidil 16) breathnaítear ar an mhórphictiúr, mar tá níos mó i gceist leis an Phlean Dhigiteach ná soláthar ar theicneolaíochtaí faoi leith. Is sprioc lárnach é earnáil láidir teicneolaíochta urlabhra agus teanga a bhunú don Ghaeilge sa tír seo, earnáil a chuirfidh teicneolaíochtaí, áiseanna agus acmhainní úrscothacha ar fáil go fadtéarmach. Tá sé ríthábhachtach go mbíonn siad seo bunaithe ar shaineolas ar struchtúr na teanga, go dtógtar iad i gcomhpháirt le pobal na Gaeilge agus ag freagairt dá gcuid riachtanas, go mbíonn siad oiriúnach don chomhthéacs sochtheangeolaíoch agus go ndíríonn siad ar na dúshláin don Ghaeilge sa lá atá inniu ann.

1.2 An ré dhigiteach

Tá teicneolaíochtaí na hurlabhra agus teanga in úsáid go forleathan sa saol atá inniu ann. Tá na meáin shóisialta lárnach i saol na n-óg agus tá grúpaí caidrimh úra á gcruthú i rith an ama; tá aistrúchán uathoibríoch ag ceadú cumarsáide idir daoine a labhraíonn teangacha éagsúla; cuardaíonn muid eolas ón “rúnaí ar an fhón”; sna hionaid glaoanna, is ríomhairí seachas daoine is minice a labhraíonn ar ais linn; san oideachas, baintear úsáid as cluichí idirghníomhacha ina mbíonn comhrá againn le carachtair fhíorúla; baineann foghlaimeoirí úsáid fhorleathan as aipeanna agus ardáin foghlama ar líne; tá áiseanna cumarsáide ag tacú leo siúd faoi mhíchumas; tá áiseanna ag teacht chun tosaigh a bhailíonn eolas ar ár nglór agus ár ngníomhaíochtaí agus a dhéanann monatóireacht ar ár sláinte coirp agus meabhair; tá forbairt á déanamh ar róbait agus ar charachtair fhíorúla mar “chomrádaithe” do sheandaoine chun cuidiú leo fanacht neamhspleách sa bhaile. Tá na teicneolaíochtaí seo i ngach áit, ach in Éirinn is trí mheán an Bhéarla amháin a fhaightear iad.

1.3 An phráinn leis an obair

Tá práinn leis an Phlean Dhigiteach, ma tá an Ghaeilge chun buntáistí na réabhlóide digití a bhaint amach, agus an “bás digiteach” a sheachaint [3]. Sa tuairisc Report on the Irish Language [12] a foilsíodh le déanaí mar chuid den 2022 European Language Equality Project [13] ba léir go raibh an Ghaeilge mar cheann de dhá theanga a bhí “ar bheagán tacaíochta nó gan aon tacaíocht” i réimse na teicneolaíochta teanga. Mar a fheicfear sa mhéid a leanas, cé go bhfuil gnéithe áirithe inar féidir dul chun cinn tapaidh a dhéanamh, ní bheidh teicneolaíochtaí coimpléascacha agus réimse leathan d’áiseanna úsáideacha ar fáil thar oíche. Tá próiseas leathan, fadtéarmach i gceist: caithfear na hacmhainní teangeolaíochta a fhorbairt mar bhunús don teicneolaíocht agus do na háiseanna a bheidh ag síolrú as. Beidh infheistíocht infreastruchtúir de dhíth – oiliúint ar thaighdeoirí leis na scileanna teanga agus teicniúla atá riachtanach. Beidh clár forbartha ag teastáil lena chinntiú go bhfuil na háiseanna a ndéantar forbairt orthu ar na cinn is riachtanaí do phobal na Gaeilge. Tá sé práinneach tabhairt faoin obair anois le nach dtiocfaidh méadú ar an bhearna teicneolaíochta idir an Ghaeilge agus an Béarla agus laghdú dá réir ar na gnéithe saoil ina mbaineann muid úsáid as an teanga.

1.4 Na dúshláin teangeolaíochta agus sochtheangeolaíochta

Beidh tosaíochtaí an Phlean Dhigitigh ag brath ar chomhthéacs teangeolaíochta agus sóisialta na Gaeilge. Is léir go bhfuil an teanga i mbaol: is ghnáth-theanga phobail í i measc cainteoirí dúchais i limistéir bheaga Ghaeltachta. Agus mar go bhfuil na Gaeltachtaí scartha amach óna chéile i gceantair imeallacha, baineann sé ó dhlús criticiúil na ndaoine ag a bhfuil an Ghaeilge mar an chéad teanga. Tá forleathnú ag teacht ar an Bhéarla sna Gaeltachtaí, go háirithe i measc aos óg digiteach na linne seo [4]. Taobh amuigh den Ghaeltacht agus go háirithe sna bailte móra, tá go leor teaghlach ag labhairt na Gaeilge sa bhaile, agus go leor daoine eile atá ag cur feabhas ar a gcumas Gaeilge agus á labhairt mar an dara teanga ina saol laethúil - ach ar an iomlán tá na Gaeilgeoirí seo scartha amach óna chéile fosta.

Ar an taobh dhearfach den scéal, tá tacaíocht Rialtais agus tacaíocht bhunreachtúil ag an Ghaeilge mar an chéad teanga náisiúnta. Aithnítear oibleagáidí an Stáit d’athbheochan na teanga sa Straitéis 20 Bliain don Ghaeilge, 2010–2030 [5]. Chun an straitéis seo a chur i bhfeidhm, leagann an Plean Gníomhaíochta don Ghaeilge, 2018 – 2022 [6] amach, ar bhealach comhtháite, na gníomhaíochtaí atá indéanta, réalaíoch agus is féidir a thomhais don tréimhse ama suas go dtí 2022. I measc na moltaí sa Phlean Gníomhaíochta tá dhá réimse a bhaineann go dlúth le saothar an Phlean Dhigitigh – Réimse 5: Na Meáin agus an Teicneolaíocht, agus Réimse 6: Foclóirí. Mar an chéad chéim eile, tá leathnú mór sa cháipéis seo ar na réimsí taighde agus forbartha a mholtar, agus tá béim ar an fhís fhadtéarmach: earnáil láidir teicneolaíochta urlabhra agus teanga don Ghaeilge a bhunú sa tír i gcomhpháirt le pobail labhartha na Gaeilge.

Ó bunaíodh an Stát, ba léir go mbeadh ról lárnach ag an chóras oideachais i seachadadh na teanga ó ghlúin go glúin. Is ábhar éigeantach scoile í go deireadh na meánscoile. Tá dúshláin mhóra ag baint le teagasc na Gaeilge, go mórmhór mar nach bhfuil teagmháil rialta ag foghlaimeoirí le cainteoirí dúchais. Tá taighde a léiríonn gur thit caighdeán na Gaeilge go mór i

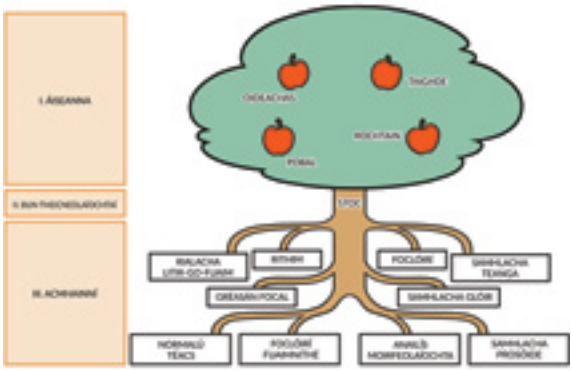
scoileanna Béarla [7]. Ach ag an am chéanna, léiríonn fás na nGaelscoileanna an rachairt atá ar scolaíocht trí mheán na Gaeilge. Tá méadú ar chúrsaí Gaeilge atá á gcur ar fáil in ollscoileanna thar lear, agus is léir ón líon daoine atá ag baint úsáid as áiseanna Gaeilge ar líne go bhfuil pobal idirnáisiúnta ag foghlaim agus ag baint úsáid as an teanga. Cuireann siad seo ar fad le neart na teanga – agus beidh ról tábhachtach ag an teicneolaíocht chun pobail scaipthe na Gaeilge a nascadh le chéile.

Tá dúshláin eile a bhaineann le gnéithe de struchtúr teangeolaíochta na Gaeilge, atá an-difriúil ón Bhéarla (féach Caibidil 2). Tá impleachtaí dá réir d'fhorbairt teicneolaíochtaí don Ghaeilge. Go dtí seo, is don Bhéarla a tógadh fórmhór na teicneolaíochta agus tá blianta fada de thaighde teangeolaíochta taobh thiar den iarracht – rud a fhágann go bhfuil an teicneolaíocht curtha in oiriúint do struchtúr an Bhéarla. Don Ghaeilge, beidh anailís eolaíoch ar struchtúr na teanga de dhíth, déanta ar bhealach a bheidh inúsáidte i dtógáil na teicneolaíochta. Cinnteoidh sé seo go mbeidh an teicneolaíocht in oiriúint don teanga, agus gur féidir an t-eolas teangeolaíochta a chur ag obair in áiseanna don phobal.

Rud eile atá tábhachtach go háirithe d'fhorbairt na dteicneolaíochtaí urlabhra: níl canúint chaighdeánach labhartha amháin ann sa Ghaeilge, ach tá trí mhórchanúint agus líon áirithe de mhionchanúintí. Tá caighdeán scríofa ann, an Lárchanúint [8] atá bunaithe ar na canúintí éagsúla, ach ní bhaineann sé le haon chanúint labhartha amháin. Mar sin, tá teicneolaíocht urlabhra ilchanúnach de dhíth don Ghaeilge – rud atá an-difriúil ó na mórtheangacha ina gcuirtear an bhéim ar an chanúint labhartha chaighdeánach. Is dúshlán ann féin go gcaithfear teicneolaíochtaí na hurlabhra (sintéis cainte, aithint cainte, córais comhrá agus na háiseanna a eascraíonn astu) a fhorbairt do na canúintí éagsúla, ach tá sé ríthábhachtach i gcomhthéacs na Gaeilge go mbeidh na háiseanna cainte a chuirfear ar fáil ag freastal ar phobal iomlán na Gaeilge.

1.5 Imlíne an Phlean

Tá an Plean roinnte i dtrí Pháirt, agus léirithe macasamhail crainn i bhFíor 1.1. Ní mór a thuiscint nach aistriú anseo is ansiúd ar theicneolaíochtaí an Bhéarla atá i gceist i bhforbairt teicneolaíochtaí urlabhra agus teanga don Ghaeilge. Chun tabhairt faoi i gceart, is próiseas fadtéarmach atá i gceist, ina mbeidh gá le buntaighde maraon le taighde feidhmeach. Mar a bheadh crann ann, chun torthaí fiúntacha a chur ar fáil (na háiseanna don phobal) is gá fréamhacha doimhne (acmhainní teangeolaíochta) agus stoc láidir (bunteicneolaíochtaí). Cé go gcuirtear síos orthu anseo ceann ar cheann, chun earraí éifeachtacha, oiriúnacha a sholáthar, tá obair chomhthreormhar de dhíth sna trí ghné agus comhoibriú eatarthu.



Figiúr 1.1 Na trí chuid den Phlean Dhigiteach

Cuid I: Is iad na fréamhacha a fheictear i bhFíor 1.1 ná na bunacmhainní teangeolaíochta agus urlabhra, atá mar bhunchloch d'fhorbairt na teicneolaíochta. Ina measc, tá corpais labhartha agus scríofa na Gaeilge, anailís dhigiteach teangeolaíochta ar strúchtúr na Gaeilge, acmhainní agus samhlacha urlabhra agus teanga. Mar atá luaite cheana, tá an cineál seo buntaighde ar siúl le blianta fada sna mórtheangacha, agus caithfear díriú air anois don Ghaeilge. Cuirfidh na hacmhainní seo lenár dtuiscint ar an teanga, agus bainfear úsáid astu i mórán áiseanna, go háirithe in áiseanna atá dírithe ar theagasc agus ar fhoghlaim na teanga.

Cuid II: Is iad na bunchórais urlabhra agus teanga stoc an chrainn, a bhaineann úsáid as na hacmhainní i gCuid 1. Is córais choimpléascacha iad seo, mar atá sintéis na cainte, aithint cainte, córais comhrá ríomhchuidithe, córais meaisínaistriúcháin, agus córais cuardaithe eolais.

Cuid III: Baineann sé seo leis na háiseanna, is é sin na torthaí i bhFíor 1.1, ina bhfuil na bunteicneolaíochtaí agus acmhainní leabaithe, agus a dhíríonn ar an phobal agus ar ghrúpaí úsáideoirí faoi leith. Is ar fhiúntas na n-áiseanna go príomha a bhraitheann leas na teicneolaíochta don phobal. Chomh maith le háiseanna ginearálta, cuirtear béim faoi leith anseo ar áiseanna do theagasc na Gaeilge agus ar áiseanna rochtana dóibh siúd faoi mhíchumas.

Nóta don léitheoir: Don ghnáthléitheoir, is fearr díriú ar an dá Chuid II & III, a mhíníonn na teicneolaíochtaí agus a léiríonn le samplaí coincréiteacha buntáistí na n-áiseanna a cheadaíonn siad. I gCuid I, tá eolas níos teicniúla ar an bhuntaighde teangeolaíochta agus ar

acmhainní atá taobh thiar den teicneolaíocht. Cé nach gá Cuid I a léamh chun ciall a bhaint as Codanna II & III, tugtar cúlra na hoibre ann agus léirmheas níos iomláine dóibh siúd ar suim leo é.

Tá gach Cuid roinnte i gcaibidlí, ina bpléitear réimsí faoi leith. Chun gur féidir leis an léitheoir súil thapa a chaitheamh thar réimsí áirithe, tá an leagan amach céanna nach mór ar fhormhór na gcaibidlí, agus baintear úsáid nuair is féidir as na fo-cheannteidil: Cén rud é? – Cé dó é agus cén tábhacht atá leis? – Cén dóigh a n-oibríonn sé? – Cad é atá déanta go dtí seo? – Cad é atá le déanamh? – Moltaí.

1.6 Nasc Idir dhá Dhisciplín: Eolaíocht na hUrlabhra & Próiseáil Teanga Nádúrtha (NLP)

Sa phlean, tá taighde comhthreomhar in dhá dhisciplín éagsúla i gceist: (i) Próiseáil Teanga Nádúrtha (NLP), a dhéanann scrúdú ar théacs, i.e., ábhar scríofa; agus (ii) Eolaíocht na hUrlabhra, a dhéanann scrúdú ar an fhuaim, i.e., an tonnchruth acústach de ghnáth. Is minic go leantar don taighde sa dá dhisciplín neamhspleách ar a chéile, ach beidh dlúthchomhoibriú ag teastáil eatarthu don Phlean Digiteach, agus go háirithe d'fhorbairt na n-áiseanna i gCuid III.

Le Próiseáil Teanga Nádúrtha (NLP) déantar anailís ríomhtheangeolaíochta ar théacsanna, is é sin ábhar atá scríofá ó thús, nó athscríofa ó ábhair labhartha. Tá leagan scríofa den teanga lárnach i saol an lae inniu, agus is ón leagan scríofa go háirithe a dhéantar anailís ar struchtúr na bhfocal, na bhfrásaí, na n-abairtí, (lena dtuigtear go minic an ghramadach) agus ar an dóigh ina gcuireann siad seo ciall in iúl. Mar sin, tá an teanga scríofa mar bhunús do mhórán shaghsanna taighde teangeolaíochta, agus don obair seo tá gá le scileanna ríomheolaíochta agus teangeolaíochta. Pléitear na gnéithe taighde seo in §2.5, agus §2.6. Tá corpais mhóra de théacsanna de dhíth fosta (Caibidil 3) mar bhunús do thógáil na n-acmhainní teanga a phléitear i gCaibidlí 4, 5 agus 6. Leis na corpais agus leis na hacmhainní seo déantar forbairt ar na teicneolaíochtaí téacs-bhunaithe (Caibidil 11 agus 12) agus ar na háiseanna a bhaineann le téacs i gCuid III.

Pléann Eolaíocht na hUrlabhra/ Próiseáil na hUrlabhra le tonnchruth acústach na fuaime de ghnáth. Tá a lán éagsúlachta sa chaint nach samhlófaí ón leagan scríofa (féach Caibidil 2.1-2.4), mar shampla na difríochtaí i bhfuaimniú na gcanúintí, na difríochtaí móra sa tonnchruth acústach ag fir, mná is páistí. Rud eile, tá mórán gnéithe de bhrí na cainte nach nochtar i dtéacs: tuigeann muid ciall na cainte, ní hamháin ó na focail (gur féidir a scríobh), ach fosta ón dóigh ina ndeirtear iad (nach féidir a scríobh). Caithfidh modhanna anailíse agus próiseála na hurlabhra na héagsúlachtaí móra acústacha sa teanga labhartha a shárú, caithfidh siad na gnéithe nach scríobhtar i mbrí na cainte a aimsiú, caithfidh siad samhail a thógáil ar na próisis lena ngineann agus lena dtuigeann muid an chaint, agus an ríomhaire a chur ag caint agus ag éisteacht mar a bheadh duine ann. Beidh dlúthchomhoibriú idir taighde sa teangeolas-foghraíochta (Caibidil §2.4) agus in innealtóireacht na hurlabhra (Caibidil 7) de dhíth, ag baint úsáid as corpais mhóra chainte (Caibidil 3) chun teicneolaíochtaí na hurlabhra a fhorbairt (Caibidil 8, 9 agus 10), maraon leis na háiseanna atá bunaithe orthu (Cuid III).

1.7 Modhanna eolasbhunaithe agus modhanna meaisínfhoghlama

Tá dhá mhodh oibre a mbainfear úsáid astu i bhforbairt na dteicneolaíochtaí atá luaite sa Phlean Dhigiteach, modh atá bunaithe ar eolas teanga agus modh a bhaineann úsáid as meaisínfhoghlaim.

Modh eolasbhunaithe: Leis an mhodh seo, déantar anailís ar rialacha struchtúir na teanga agus baintear úsáid astu le córais teicneolaíochta a fhorbairt. Mar shampla, don tsintéis, is féidir úsáid a bhaint as ár n-eolas ar na rialacha a aistríonn ó litir-go-fuaim, as foclóirí fuaimnithe, as múnlaí prosóide agus as rialacha na morfóineolaíochta. Baineann na caibidlí i gCuid I den phlean go mhórmhór le soláthar an eolas seo i riocht atá inúsáidte don teicneolaíocht.

Modh staitisticiúil meaisínfhoghlama: Leis an mhodh seo, baintear úsáid as corpais ollmhóra, mar shampla, dhá chorpas leis an téacs chéanna in dhá theanga dhifriúla, agus aimsíonn feidhmchláir staitisticiúla na comhfhreagairtí eatarthu, chun córas ríomhaistriúcháin a thógáil. Tá na modhanna meaisínfhoghlama tagtha chun cinn go mór le blianta beaga anuas le forbairt ar Líonraí Néaracha Doimhne (Deep Neural Nets, DNN). I ngnéithe áirithe teicneolaíochta, is féidir dul chun cinn an-tapa a dhéanamh nuair atá corpais ollmhóra oiriúnacha ar fáil. Tá na modhanna meaisínfhoghlama seo lárnach don aistriú uathoibríoch agus don aithint cainte (Caibidil 9, 11).

I bhforbairt na teicneolaíochta, beidh an dá mhodh in úsáid, agus buntáistí faoi leith ag gach ceann acu. Is sprioc thábhachtach ann féin go mbeadh tuiscint dhomhain agus mionchuntas digiteach againn ar struchtúr na Gaeilge mar a labhraítear inniu í. Nuair a thógtar córas ar bhunús eolais teangeolaíochta, is féidir feabhas a chur air de réir mar a mhéadaíonn ár dtuiscint

ar an teanga. Agus mar go bhfuil an t-eolas agus na hacmhainní i bhfoirm dhigiteach, is féidir iad a athúsáid in iomaí áis: tá géarghá leo go háirithe in áiseanna do theagasc na teanga. Amach anseo, beidh ardáin foghlama 'chliste' d'fhoghlaimeoirí ag brath, ní hamháin ar na teicneolaíochtaí féin, ach fosta, ar na hacmhainní teangeolaíochta taobh thiar díobh. Tá bunús eolais (i bhfoirm dhigiteach) tábhachtach ar a lán bealaí don fhorbairt fhadtéarmach, ach is é an míbhuntáiste leis an mhodh oibre seo ná go mbíonn go leor ama i gceist agus ardscileanna teangeolaíochta de dhíth.

Leis na modhanna staitistiúla measínfhoghlama, is féidir gnéithe áirithe forbartha a chur i gcrích go tapa gan dua. Tá an-dul chun cinn déanta sa Bhéarla i gcórais bunaithe ar Líonraí Néaracha Doimhne. Táthar á n-úsáid fosta go héifeachtach faoi láthair i bhforbairt na teicneolaíochta don Ghaeilge agus is léir go mbeidh siad an-tábhachtach sa todhchaí. Ach mar sin féin, ní hionann an cás sa dá theanga. Tá corpais ollmhóra ar fáil sa Bhéarla agus diantaighde teangeolaíochta ar siúl le fada an lá, rud a cheadaíonn úsáid sháréifeachtach as modhanna measínfhoghlama. Is míbhuntáiste le córais bunaithe ar mheaisínfhoghlaim, nach bhfaightear aon tuiscint astu ar an dóigh ina n-oibríonn siad nó aon tuiscint breise ar struchtúr na teanga féin, gur féidir a úsáid (mar shampla) in áiseanna foghlama. Agus mar a thagann méadú ar ár n-eolas teangeolaíochta, ní féidir uasdátú ná feabhsú a dhéanamh dá réir ar an chóras.

Beidh an dá mhodh forbartha i gceist sa Phlean Dhigiteach, agus beidh siad in úsáid le chéile d'fhorbairt córais ar leith.

1.8 Tiomantas fadtéarmach

Cé gur mar shraith caibidlí a bhaineann le réimsí éagsúla atá an cháipéis seo leagtha amach, ní ceart breathnú ar an Phlean Dhigiteach mar liosta de tháirgí le soláthar. Macasamhail an chrainn i bhFíor 1.1, braitheann na codanna éagsúla ar a chéile, agus ba cheart go leanfadh sé ag fás agus go mbeadh torthaí úra ag síolrú as. Agus fiú nuair atá córas teicneolaíochta nó áis faoi leith tógtha, ní sin deireadh an scéil. Beidh feabhsú leanúnach de dhíth, fadhbanna le réiteach, tástáil agus measúnú le déanamh, le cinntiú go bhfuil sé seasmhach agus in oiriúint do na húsáideoirí. Agus de réir mar a thagann borradh ar earnáil na teicneolaíochta Gaeilge, beidh córais agus áiseanna nua ag teacht chun tosaigh ina mbeidh gnéithe difriúla an taighde fite fuaite. Is plean fadtéarmach atá i gceist agus tá sé mar sprioc aige go mbeidh an Ghaeilge leabaithe go daingean sa teicneolaíocht de réir mar a thagann sí chun cinn. Tá na moltaí a leanas dírithe ar an sprioc seo, agus tá plé níos iomláine orthu sa Chonclúid (Caibidil 16).

Infrastruchtúr a bhunú le hionaid barr fheabhais le foirne idirdhisciplíneacha agus le hardleibhéal Gaeilge: tá scileanna ag teastáil i réimsí éagsúla, mar a bhfuil teangeolas/foghraíocht, ríomheolaíocht, innealtóireacht na hurlabhra - agus ardleibhéal Gaeilge. Is é an deacracht is mó ag na foirne taighde go dtí seo ná teacht ar thaighdeoirí oilte a bhfuil an leibhéal cuí Gaeilge acu. Don todhchaí, tá sé ríthábhachtach go mbeidh muid ag cur oiliúna ar thaighdeoirí leis na sainscileanna cuí agus ardchumas teanga. Tá cúrsaí oiliúna ollscoile ann a dhéanann freastal go pointe áirithe ar an tsaghas oiliúna idirdhisciplíní atá de dhíth, agus moltar iad a leathnú agus a neartú chun freastal mar is cóir ar theangeolaíocht na Gaeilge, agus chun freagairt ar bhealach níos iomláine ar réimsí taighde an Phlean Dhigitigh. Beidh feachtas de dhíth fosta chun mic léinn chumasacha le hardleibhéal Gaeilge a mhealladh isteach san earnáil.

Tógáil ar an obair atá ar siúl faoi láthair: Cé go bhfuil píosa fada le dul chun go mbeidh réimse leathan teicneolaíochta urlabhra agus teanga ann don Ghaeilge [2], tá tréan oibre ar siúl i réimsí áirithe faoi láthair, mar a léirítear i gcuid de na caibidlí. Beidh sé tábhachtach tógáil ar an dúshraith atá ann faoi láthair agus réimsí agus scála an taighde agus na hoiliúna a leathnú amach chun gur féidir spriocanna an Phlean a bhaint amach.
Buntaighde agus taighde feidhmeach: Tá meascán de bhuntaighde agus taighde feidhmeach de dhíth (ó fréamhacha go torthaí i bhFíor 1.1). Is bunchloch don teicneolaíocht an buntaighde agus tá sé tábhachtach do thodhchaí na hoibre. Tá an taighde feidhmeach lárnach chun aschur an taighde a aistriú ina n-áiseanna a bheidh ag freagairt do riachtanais an phobail.

Leanúnachas: Tá leanúnachas maoinithe níos tábhachtaí fiú ná leibhéal an mhaoinithe a chuirtear ar fáil. Gan leanúnachas, titeann foirne taighde as a chéile, mar go mbíonn ar thaighdeoirí díriú ar fhostaíocht fhadtéarmach, agus tá an baol ann go gcaillfear an infheistíocht sa taighde agus in oiliúint ar thaighdeoirí. Mar sin, beidh leanúnachas sa mhaoiniú de dhíth do na hionaid taighde chun gur féidir leo fás de réir mar a fhásann an teicneolaíocht. Beidh pleanáil fhadtéarmach de dhíth dó seo agus comhthéacs le bunú ina bhfeiceann taighdeoirí óga go bhfuil gairm bheatha fhiúntach le baint amach i dtaighde agus i bhforbairt teicneolaíochta na Gaeilge.

Comhpháirtíocht le Pobal na Gaeilge: braitheann éifeachtacht an Phlean ar leibhéal na comhpháirtíochta le pobal na Gaeilge. Luaitear pointí áirithe in §1.10 thíos, agus tá plé níos iomláine sa Chonclúid, Caibidil 16 ar bhealaí ina bhféadfaí an nasc leis an phobal a chothú.

1.9 Impleachtaí do mhionteangacha agus do theangacha eile i mbaol

Tuigtear anois chomh tábhachtach is atá teicneolaíochtaí na hurlabhra agus teanga i slánú mionteangacha agus teangacha i mbaol. Gan iad, éiríonn réimse úsáide na teanga níos imeallaí fós, rud a chuireann leis an luas scanrúil lena bhfuil teangacha á gcailliúint sa lá atá inniu ann. Tá tuiscint ag fás i measc taighdeoirí sa réimse nach fada go mbeidh sé rómhall an treocht a mhaolú agus gur mithid beart a dhéanamh anois. Tá sé seo le feiceáil chomh maith i leithéidí an Erasmus+ Digital Language Diversity Project [10], an SIGUL Special Interest Group for Under-resourced Languages [11], an tuairisc Language Equality Report [1] ar glacadh leis ag leibhéal an Aontais Eorpaigh le déanaí, agus an tionscadal European Language Equality [13].

Tá sé mar bunphrionsabal ag an Phlean Dhigiteach, díriú ar na dúshláin atá sonrach don Ghaeilge agus tús áite a thabhairt do na gníomhaíochtaí is práinní do lucht labhartha na teanga. Ar an ábhar seo, ní mór don phlé a bheith mionsonraithe agus bunaithe ar an chomhthéacs áitiúil. Ach ag an am chéanna, is minic na dúshláin agus fadhbanna céanna ag teangacha i mbaol. Fágann sin go bhfuil impleachtaí ag eispéireas na hÉireann do chásanna eile: beidh na modhanna oibre – na bealaí ina dtugann muid faoi na dúshláin agus na bealaí ina dtagann muid ar réiteach don Ghaeilge – thar a bheith luachmhar do phobail eile atá ag iarraidh a dteangacha a chaomhnú.

Mar mhionteanga, tá stádas faoi leith ag an Ghaeilge, sa mhéid is go bhfuil aitheantas agus tacaíocht aici ón Rialtas, mar a fheictear i dtionscnamh an Phlean Dhigitigh. Le feidhmiú an Phlean is féidir linne tacú go gníomhach le pobail mionteangacha eile. Cheana féin, tá grúpaí sa tír seo ag tacú le grúpaí thar lear nach bhfuil na féidearthachtaí céanna acu. Tá sé mar chuspóir ag an Phlean Dhigiteach, go roinnfear ár dtaithí, ár saineolas agus nuair is cuí, ár n-acmhainní, le pobail teangacha eile i mbaol.

1.10 An nasc leis an Ghaeltacht agus le Pobal na Gaeilge

Tá sé tábhachtach nach mbeidh an Plean Digiteach scartha amach ó phobal na Gaeilge, agus go háirithe ó chainteoirí dúchais na Gaeltachta. Tá comhpháirtíocht leis an phobal de dhíth i ngach gné – chun infrastruchtúr a thógáil, oiliúint a chur ar thaighdeoirí, cinntí a dhéanamh maidir le tosaíochtaí na dteicneolaíochtaí, áiseanna a scaipeadh i measc an phobail. Mar atá luaite thuas, beidh gá le hoiliúint idirdhisciplíneach a chur ar thaighdeoirí na todhchaí, chun go mbeidh scileanna teicniúla, teangeolas agus ardchumas teanga acu. Beidh cainteoirí dúchais na Gaeltachta agus daoine eile le hardchumas Gaeilge lárnach don fheachtas seo.

Mar a théann an Plean in aibíocht, beidh deiseanna ann le gnólachtaí campais a bhunú a thabharfadh áiseanna ar leith don mhargadh. Is fiú pleanáil chuí a dhéanamh dá leithéid ó thús. Beidh polasaí le forbairt: sa mhéid is gur le hairgead poiblí a dhéantar forbairt ar an teicneolaíocht, beidh treoirlínte de dhíth i dtaca leis an sciar d'aon bhrabús a théann chun tacú le taighde an Phlean.

I gcás gnólachtaí campais, bheadh lonnú Gaeltachta (nó comhlonnú idir Gaeltacht agus Ollscoil/Ionad Taighde) an-oiriúnach, tairbheach don tionscnamh, agus thacódh sé le ról phobail na Gaeltachta i gcaomhnú na teanga don chéad glúin eile. Tá plé níos iomláine ar na pointí seo sa Chonclúid, Caibidil 16.

1.11 Ár n-Oidhreacht sa Ré Dhigiteach

Le maoiniú réasúnta, le dea-stiúradh agus le súil ar an fhís fhadtéarmach, tá féidearthachtaí ag an Phlean Dhigiteach uirlisí cumhachtacha a chur i lámha phobal labhartha na Gaeilge, a thacóidh leo in athbheochan, caomhnú, agus doiciméadú na teanga. Is í an teanga féin an tseoid is luachmhaire cultúrtha againn agus is tríd an teanga atá teacht againn ar na mórsheoda eile. Gan an teanga, caillfear cheann de na traidisiúin liteartha is ársa san Eoraip. Gan an teanga, níl teacht againn ar an taisce mhór d'amhráin, d'fhilíocht, de bhéaloideas is d'fhinscéalta – a théann siar na céadta bliain, agus a cheanglaíonn lenár n-áit dúchais sinn. Le breacadh na ré digití don Ghaeilge, tá sé mar chuspóir ár n-oidhreacht a nascadh le hÉirinn na nua-aoise, rud a chuideoidh linn greim a choinneáil ar ár bhféiniúlacht i measc an chultúir leamh, aonchineálaigh dhomhanda.

Tagairtí

[1] Evans, J., Report on Language Equality in the Digital Age, léite 23 Samhain 2019 ar http://www.europarl.europa.eu/doceo/document/A-8-2018-0228_EN.html

[2] Ní Chasaide, A., Ní Chiaráin, N., Berthelsen, H., Wendler, C. agus Murphy, A. (2015). Speech technology as documentation for endangered language preservation: the case of Irish. Proceedings of the 18th International Congress of Phonetic Sciences, Glaschú

[3] Judge, J., Ní Chasaide, A., Ní Dhubhda, R., Scannell, K. agus Uí Dhonnchadha, E. (2012) An Ghaeilge sa Ré Dhigiteach. In META-NET Language White Paper Series Europe's Languages in the Digital Age. Springer.

[4] Ó Giollagáin, Conchúr agus Charlton, Martin (2015) Nuashonrú ar an Staidéar Cuimsitheach Teangeolaíoch ar Úsáid na Gaeilge sa Ghaeltacht 2006–2011: Príomhfhaisnéis na Limistéar Pleanála Teanga. Údarás na Gaeltachta, Éire, 2015

[5] An Straitéis 20 Bliain don Ghaeilge 2010–2030. An Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meán

[6] An Plean Gníomhaíochta don Ghaeilge, 2018-2022 An Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meán

[7] Harris, J. (2006). Irish in Primary Schools: Long-Term National Trends in Achievement. Baile Átha Cliath: An Roinn Oideachais agus Eolaíochta

[8] Ó Baoill, D. (1986). Lárchanúint don Ghaeilge, Iml. 14, Tuarascail Taighde, Institiúid Teangeolaíochta Éireann, Baile Átha Cliath.

[9] Goodfellow, I., Bengio, Y. agus Courville, A. (2016) Deep Learning. MIT Press.

[10] The Digital Language Diversity Project Erasmus+ Programme | Comhaontú Deontais uimh. 2015-1-IT02-KA204-015090, www.dldp.eu/en

[11] SIGUL: Special Interest Group: Under-resourced Languages, a joint special-interest group of the European Language Resources Association (ELRA) and of the International Speech Communication Association (ISCA), www.elra.info/en/sig/sigul/

[12] Lynn, Teresa (2022). Doiciméad insoláthartha D1.20 Report on the Irish Language. Tionscal insoláthartha; EU project European Language Equality (ELE); Comhaontú Deontais uimh. LC- 01641480 – 101018166 ELE. URL: https://european-language-equality.eu/wp- content / uploads / 2022 / 03 / ELE _ _ _ Deliverable _ D1 _ 20 _ _Language _ Report _ Irish_.pdf.

[13] Ar fáil: https://european-language-equality.eu/

Caibidil 2
Anailís Teangeolaíochta agus Doiciméadú Digiteach ar an Ghaeilge Labhartha agus Scríofa

2.1 Cad é atá i gceist le hanailís teangeolaíochta?

Is éard atá i gceist le hanailís teangeolaíochta ná taighde eolaíoch ar struchtúir na teanga, i.e. an cód trína gcuireann muid ár smaointe in iúl, agus a cheadaíonn an chumarsáid eadrainn. Tá anailís fhollasach ar struchtúir na Gaeilge de dhíth chun teicneolaíochaí éifeachtacha a fhorbairt a bheidh ag freagairt do riachtanais an phobail. Don Phlean Digiteach, is gá modhanna anailíse a úsáid a chuireann an fheasacht teangeolaíochta ar fáil i riocht is féidir a chur i bhfeidhm ar thógáil na teicneolaíochta urlabhra agus teanga. Moltar fosta go gcuirfear an t-eolas ar struchtúr na teanga ar fáil ar bhealaí is féidir a úsáid in áiseanna do theagasc na teanga, in áiseanna rochtana dóibh siúd faoi mhíchumas agus in áiseanna don phobal i gcoitinne.

Tugtar léiriú sa chaibidil seo ar na gnéithe éagsúla de struchtúir na teanga a mbeifear ag díriú orthu, agus ar an mhéid atá déanta go dtí seo. Tugann sé cúlra na gcaibidlí a leanas i gCuid I, agus an comhthéacs don fhorbairt teicneolaíochta i gCodanna II agus III.

Sa teanga labhartha agus scríofa baintear úsáid as dhá mheán atá an-difriúl, an fhuaim agus an téacs, agus mar atá luaite i gCaibidil 1.6, tá dhá réimse taighde dá réir sa Phlean Dhigiteach. Tá réimse amháin téacsbhunaithe (ag brath ar fhoinsí scríofa, nó fiú ar leaganacha tras-scríofa de chorpais cainte), agus baintear úsáid as teicnící próiseála teanga nádúrtha (PTN). Tá an dara réimse bunaithe ar thaifead den chaint féin (m.sh. tonnchruth na fuaime), agus baintear úsáid as modhanna anailíse agus próiseála na urlabhra.

Nuair a labhraíonn muid, tá struchtúrú ar shruth na fuaime i ngach teanga a cheadaíonn don éisteoir ciall a bhaint as an tsraith fuaimeanna, as na focail agus as na habairtí. Nuair a scríobhann muid, baineann muid úsáid as litreacha chun sraith na bhfuaimeanna a léiriú, spás idir focail chun tús/deireadh focail a aimsiú, camóg, lán stad agus cinnlitreacha chun tús/deireadh frásaí agus abairtí a aimsiú. Ach tá difríochtaí tábhachtacha idir an teanga scríofa agus an teanga labhartha, agus impleachtaí dá réir don bhealach ina ndéantar forbairt ar theicneolaíochtaí urlabhra (caintbhunaithe) agus teanga (téacsbhunaithe).

Tugann an leagan scríofa foirm ‘idéalach’ den teanga dúinn, ina bhfágtar ar lár na stadanna agus na ‘botúin’ chainte den leagan labhartha. Baintear úsáid fhorleathan as téacs, go háirithe i gcumarsáid ríomhchuidithe, agus tá an teanga scríofa tábhachtach san oideachas, i gcórais dlí agus riaracháin. Tá téacs seasmhach, éasca le scaipeadh agus is féidir é a stóráil go fadtéarmach i gcartlanna. Éascaíonn an leagan scríofa go mór an anailís ar struchtúir na bhfocal, na bhfrásaí agus na n-abairtí – na gnéithe struchtúracha a thuigtear go minic mar ghramadach na teanga. I mórán teangacha tá stór mór de théacsanna ar fáil – rud atá fíor-úsáideach do thógáil na teicneolaíochta. Ach chun múnlú ríomhaire a dhéanamh ar na gnéithe struchtúracha seo agus chun leas a bhaint as teicnící nuálacha cumhachtacha, tá níos míos mó ná stór téacsanna de dhíth. Tá gá fosta le heolas ar struchtúir (ghramadaí) na bhfocal, na bhfrásaí, na n-abairtí, na n-alt agus ar an dóigh ina n-oibríonn siad le chéile chun ciall na teanga a léiriú. Tá tuiscint ó dhúchas ag gach duine ar chód (struchtúr gramadaí) a dteanga féin, ach chun é seo a chur ar fáil don ríomhaire tá gá le hanótáil agus le meiteashonraí . In §2.5 agus §2.6 thíos, pléitear na gnéithe seo de struchtúir na teanga, mar chúlra do Chaibidlí 3, 4, agus 5 ina bhfuil cur síos ar na stóir téacsbhunaithe (corpais, bunachair agus líonraí sonraí) chomh maith leis na teicnící agus na huirlisí don phróiseáil teanga nádúrtha (PTN) atá in úsáid i bhforbairt na teicneolaíochta teanga.

Is modhanna oibre difriúla atá i gceist le taighde ar an teanga labhartha, agus ar thonncruth na fuaime. Sa chaint, níl na fuaimeanna ná na focail discréideach, seasmhach ná scartha amach óna chéile mar atá sa scríobh: athraíonn siad mar a athraíonn comhthéacs na bhfuaimeanna mórthimpeall, an té atá ag labhairt, an chanúint srl. Caithfidh teicneolaíocht na hurlabhra déileáil leis an tsíormhalartú seo. Rud tábhachtach eile: tá a lán gnéithe eolais sa chaint nach bhfuil nótáil orthu sa leagan scríofa. Tuigeann muid brí na cainte, ní hamháin ó na focail a deirtear (gur féidir a scríobh) ach fosta ón dóigh ina ndeirtear iad – ón phrosóid (nach scríobhtar). Fiú le frása simplí mar "Tá mé", ag brath ar an dóigh ina ndeirtear é, athraíonn brí na bhfocal go hiomlán – le ciall dhearfach, dhiúltach, mar abairt shimplí, mar cheist, mar ordú. Nochtann cáilíochtaí an ghutha ár seasamh i dtaca leis an éisteoir de (formálta, cairdiúil); ár mothúcháin (feargach, brónach) agus ár meon i dtaca le comhthéacs na cainte (nochtann muid suim nó easpa suime). Is ó mhodhnú leanúnach an ghlóir a thagann beocht agus cuid mhaith de bhrí na cainte – agus is gnéithe tábhachtacha iad seo do theicneolaíocht na hurlabhra. Chun struchtúr na teanga labhartha a scrúdú do theicneolaíocht labhartha na Gaeilge, tá taighde idirdhisplíneach de dhíth, ina bhfuil lucht foghraíocht-teangeolaíochta (§2.4) agus lucht innealtóireachta na hurlabhra (Caibidil 7) ag obair as lámha a chéile, ag baint úsáid as corpais labhartha (Caibidil 3) atá deartha go sonrach do riachtanais na dteicneolaíochtaí éagsúla urlabhra (Caibidlí 8-10).

Sa taighde ar an teanga labhartha (agus ar theicneolaíocht na hurlabhra don Ghaeilge) caithfear cur san áireamh nach leagan caighdeánach amháin atá i gceist (mar is amhlaidh don teanga scríofa) ach trí mhórchanúint chomh maith le fochanúintí, atá uilig ar aon chéim i dtaobh an chaighdeáin de. Sna mórtheangacha, bhí an bhéim ar fhorbairt na teicneolaíochta don chanúint chaighdeánach, is é sin canúint na huasaicme ó thaobh na staire de. Sa Ghaeilge, caithfear díriú ón tús ar na canúintí uile.

In ainneoin na ndifríochtaí idir spriocanna, scileanna agus teicnicí taighde don teanga labhartha agus an teanga scríofa, is fiú a lua go mbíonn aschur taighde ón dá réimse fite fuaite le chéile i mórán de na teicneolaíochtaí i gCuid II agus III den Plean Dhigiteach.

Sna míreanna a leanas, pléitear tábhacht an taighde in §2.2 agus tugann §2.3 léargas gairid ar na gnéithe éagsúla teangeolaíochta le fiosrú. Dóibh siúd ar mian leo tuilleadh eolais, tá cur síos níos iomláine ar na gnéithe seo i míreanna §2.4-§2.6. Tugtar cuntas in §2.7 ar sealbhú teanga agus in §2.8 ar mhalartú teanga, réimsí taighde atá tábhachtach fosta don Phlean Dhigiteach.

2.2 Cén gá atá leis?

Tá taighde teangeolaíochta mar bhunús do na hacmhainní teanga agus urlabhra i gCaibidlí 3-7, atá riachtanach do thógáil na mbunteicneolaíochtaí i gCuid II den Phlean. Tá sé tábhachtach fosta go mbíonn feasacht ar struchtúir na Gaeilge ag lucht forbartha na teicneolaíochtaí, chun a chinntiú go bhfuil an teicneolaíocht curtha iomlán in oiriúint don teanga. Tá níos mó i gceist le forbairt na teicneolaíochta Gaeilge ná córas atá cheana-forbartha (do theanga eile) a ghlacadh mar atá agus mionathruithe imeallacha a dhéanamh dó.

Tá feasacht teangeolaíochta lárnach fosta do mhórán de na háiseanna i gCuid III den Phlean, trína bhfeictear an tairbhe do phobal na Gaeilge. Tá réimse agus cumhacht na n-áiseanna seo ag fás ó lá go lá, m.sh. áiseanna don phobal mar uirlisí cuimsitheacha profála, ceangailte le féidearthachtaí aistriúcháin, gléasanna cumarsáide agus cúntóirí digiteacha pearsanta (Caibidil 13). Tá tábhacht ar leith le háiseanna do theagasc na Gaeilge, mar gur féidir feasacht ar struchtúir na teanga a chur i bhfeidhm ar bhealach díreach i gcórais chliste FRT (foghlama ríomhchuidithe teanga, Caibidil 14). Má nasctar na teicneolaíochtaí cumhachtacha le saineolas ar struchtúir (§2.4 - §2.5) agus sealbhú (§2.7) na teanga, d'fhéadfadh muid athruithe réabhlóideacha a fheiceáil i dteagasc na teanga. Tá an-tábhacht ag baint fosta le háiseanna don rochtain agus míchumas, chun freastal i gceart orthu siúd le fadhbanna teanga, urlabhra agus litearthachta (Caibidil 15 agus §2.6.3) agus chun ceadú dóibh siúd faoi mhíchumas páirt iomlán a ghlacadh i saol na Gaeilge.

Tá sé mar aidhm bunúsach ag an Phlean Dhigiteach doiciméadú iomlán, digiteach a chur ar fáil ar struchtúir na Gaeilge. Méadófar ár dtuiscint ar an teanga, díreofar ar ghnéithe struchtúra nach ndearnadh anailís go fóill orthu, gheofar saineolas úr ar thréithe suntasacha na gcanúintí éagsúla. Spreagfar taighdeoirí agus cuirfear acmhainní luachmhara agus modhanna anailíse nuálacha ar fáil do scoláirí na todhchaí.

2.3 Cad é mar a léirítear ciall i struchtúr na teanga?

Tugtar cuntas gairid anseo ar na bealaí a léirítear brí na teanga trí ghnéithe de struchtúir teanga (réimsí teangeolais), agus tá cur síos níos iomláine ar na réimsí seo sna míreanna a leanas, §2.4-§2.6 .

Tá struchtúr na teanga casta, agus bíonn próisis éagsúla ag obair i dteannta a chéile. Mar shampla, nuair a éisteann muid le sruth na cainte, baineann muid ciall as de bharr dhá leibhéal struchtúir. Sa chéad áit tá sraith d'fhuaimeanna (consain agus gutaí) ina n-aimsíonn muid focail, frásaí,

abairtí. Neamhspléach air seo, tá bunfhuaim an ghlóir féin, as a gcruthaítear na consain agus gutaí, ag athrú de shíor le patrúin tuiníochta, cáilíochtaí an ghutha, malartú béime agus rithime – tugtar an phrosóid air seo. Ligeann an phrosóid dúinn díchódú a dhéanamh ar shruth na bhfuaimeanna, agus na haonaid mar fhocail, frásaí agus abairtí a aimsiú. Nochtann an phrosóid fosta mothúcháin agus meon an chainteora, gné de bhrí na cainte nach bhfuil teacht air sa leagan scríofa. Is iad an Fhoghraíocht agus an Fhóineolaíocht na réimsí teangeolaíochta a scrúdaíonn struchtúr na bhfuaimeanna agus na prosóide agus tá cur síos orthu in §2.4.

Nascann fuaimeanna le chéile i bhfocail. Is minic go bhfuil an focal féin curtha le chéile as mionpháirteanna (moirféimí) a bhfuil brí dá gcuid féin acu: mar shampla na trí mhionpháirt san fhocal míádhúil: mí-ádh-úil. Tá dhá shaghas brí i gceist i bhfocal: tá brí inmheánach focail, mar a bhfaighfeá i bhfoclóir, agus brí ghramadúil, a nochtann comhfhreagras an fhocail le focail eile na habairte. Nochtann moirféamanna an fhocail gnéithe éagsúla de bhrí atá leis: tá brí inmheánach iomlán difriúil ag ádh agus mí-ádh; tá brí ghramadiúil difriúl ag ádh (ainmfhocal) agus ádh-úil (aidíocht). I gcás na Gaeilge, is í an Mhoirfeólaíocht (deilbhíocht na bhfocal) is mó a phléitear sna cuntais ghramadaí thraidisiúnta, agus tá cur síos ar an réimse seo in §2.5.1.

Téann grúpaí focail le chéile i bhfrasaí agus in abairtí. Is í an Chomhréir (§2.5.2) an réimse teangeolaíochta a scrúdaíonn an dóigh a gcuireann focail agus frásaí na habairte brí ghramadúil na habairte in iúl. Braitheann brí ghramadúil na n-abairtí ar an chomhfhreagras ghramadúil idir na focail sa fhrása (ag brath ar mhionpháirteanna an fhocail agus ar ord na bhfocal sa fhrása), agus idir na frásaí san abairt.

Tá dlúthbhaint idir an chomhréir agus réimse na Séimeantaice (Mír 2.5.3) ina ndírítear ar idirghabháil bhrí inmheánach na bhfocal le comhfhreagras gramadúil na bhfocal agus na bhfrásaí san abairt. Ach mar sin féin, i ndeireadh na dála tagann an chiall a bhaineann muid as dreas léitheoireachta nó cainte – ní hamháin ó chiall inmheánach na bhfocal agus ón chomhfhreagras eatarthu– ach fosta ón chomhthéacs agus (i gcás na cainte) ón dóigh ina ndeirtear iad. Is comhthéacs leathan atá i gceist, ina gcuirtear san áireamh an chaint/téacs atá tagtha roimhe, an seasamh idir an cainteoir agus an t-éisteoir (nó idir an scríbhneoir agus an léitheoir), meon agus mothúcháin an chainteora (nó an scríbhneora): cuireann siad seo uilig lenár dtuiscint ar chiall na habairte. Pléann réimse na Pragmataice (§2.6.2) na gnéithe iomlána de bhrí na teanga.

Sna míreanna a leanas, §2.4-2.6, tá cuntas níos iomláine ar na réimsí teangeolaíochta seo, ina mínítear an méid atá bainte amach go dtí seo, agus an ceangal le tógáil na n-acmhainní i gCuid I agus na teicneolaíochtaí i gCodanna II agus III. Don léitheoir a théann díreach go dtí na codanna a bhaineann leis an teicneolaíocht, is féidir féachaint siar ar na míreanna seo nuair atá míniú ar ghné faoi leith de dhíth.

2.4 Mar a chuireann struchtúrú na fuaime brí na cainte in iúl

Is beag tuiscint ag an ghnáthdhuine ar struchtúir fuaime na teanga, i gcomparáid leis an tuiscint ar struchtúr na gramadaí – struchtúr na bhfocal, na bhfrásaí, na n-abairtí (§2.5, §2.6), atá níos soiléire sa teanga scríofa. Déantar faillí go háirithe ar an ghné seo i dteagasc na Gaeilge, agus is minic gurb í an ghné is laige ag foghlaimeoirí í. Is réimse í dá bhfuil impleachtaí móra ag an teicneolaíocht.

Tugtar an Fhoghraíocht agus an Fhoinéolaíocht ar an dá réimse teangeolaíochta a bhaineann le hanailís na fuaime chun na patrúin struchtúracha a aimsiú a cheadaíonn don éisteoir teacht ar bhrí na cainte.

Cé go bhféachtar go minic orthu siúd mar réimsí neamhspleácha, i bhformhór taighde na Gaeilge pléitear mar dhá ghné den ábhar amháin iad. Mar an chéanna sa cháipéis seo: baintear úsáid as an dá théarma gan idirdhealú eatarthu. In amanna baintear úsáid as an téarma foghraíocht-teangeolaíochta, mar nod don dá thaobh.

2.4.1 Dhá phróiseas i ngininúint na cainte: dhá shruth eolais

Chun an nasc idir struchtúir na fuaime agus brí na cainte a léiriú, ní mór a thuiscint go bhfuil dhá phróiseas neamhspleácha i gceist i nginiúint na cainte . Eascraíonn patrúin fuaime ar leith astú araon, agus léiríonn gach ceann acu gnéithe éagsúla de bhrí na cainte.

(a) Giniúint na fuaime: an glór-fhoinse (agus an phrosóid): cuireann sruth na hanála na téada gutha ag luascadh, rud a ghineann tonnchruth acústaice – an glór-fhoinse. Le linn cainte, bíonn an tonnchruth seo ag athrú de shíor, rud a chruthaíonn patrúin tuiníochta agus cáillíochta an ghutha, patrúin bhéime na siollaí agus rithim na cainte (tugtar an phrosóid ar na patrúin seo).

(b) Scagadh na fuaime: an urlabhraíocht: tá an béal mar a bheadh scagaire fuaime ann, a dhéanann

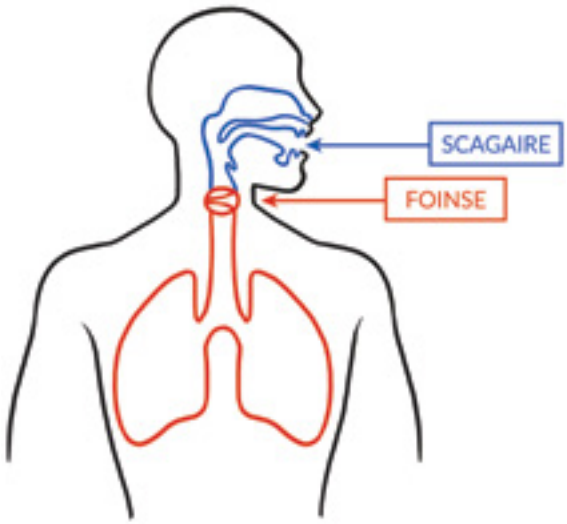
maolú ar thréithe acústaice thonnchruth an ghlóir, ar a bhealach ó na téada gutha go dtí na beola. Le linn cainte, athraíonn cruth an scagaire le bogadh na mball urlabhra (an urlabhraíocht). Le hathruithe an scagaire agus tréithe acústaice na fuaime, déantar idirdhealú ar na gutaí agus ar na consain éagsúla sa teanga. Ní athraíonn an scagaire patrúin prosóide an glór.

Mar sin, tá an glór mar fhoinse fuaime ag gutaí agus consain, agus is trí scagadh an bhéil a dhéantar idirdhealú eatarthu. Baineann an t-éisteoir brí na bhfocal féin as an tsraith thapa de ghutaí agus consain (§2.4.2), agus brí níos leithne na habairte as na patrúin phrosóide (§2.4.3).

Is le hanailís cluaise a chuirtí síos go traidisiúnta ar ghnéithe d'fhuaimniú na teanga, agus baintear úsáid as comhghnásanna idirnáisiúnta [1] le trascríobh a dhéanamh ar fhuaimeanna (gutaí/consain), agus ar phatrúin prosóide teanga, m.sh.an cuntas in [2, 3] ar fhuaimeanna na Gaeilge. Ceadaíonn modhanna nua-aimseartha digiteacha dúinn anailís eimpíreach a dhéanamh ar thréithe acústaice agus urlabhraíochta. Feictear san anailís seo nach bhfuil fuaimeanna na teanga discréideach, seasmhach ná scartha óna chéile (mar a shamhlófaí ó na litreacha scríofa). Tá an-tionchar ag fuaim amháin ar na fuaimeanna roimhis agus ina dhiaidh, agus tá an-tionchar fosta ag an chomhthéacs prosóide ina bhfaightear an fhuaim. Chomh maith leis sin, braitheann na tréithe acústaice ar mhéid bhaill urlabhra agus théada gutha an chainteora: mar sin, tá difríochtaí móra sa tonnchruth fuaime ó pháistí (ag aoiseanna difriúla) ó mhná agus ó fhir. Ní fadhb ar bith é seo don éisteoir daonna, ach ní hionann don ríomhaire: caitear na gnéithe cainte seo uilig a chur san áireamh nuair atá córais sintéise (Caibidil 8) agus aitheanta cainte (Caibidil 9) á dtógáil. Rud eile, atá luaite cheana: don Ghaeilge tá difríochtaí idir na canúintí, agus impleachtaí dá réir don taighde foghraíocht-teangeolaíochta agus d'fhorbairt theicneolaíocht na hurlabhra.

Don Phlean Dhigiteach beidh modhanna anailíse de dhíth óna bhfaightear aschur taighde atá inúsáidte sa teicneolaíocht. Dó seo, beidh dlúthchomhoibriú de dhíth idir an taighde foghraíocht-teangeolaíochta agus an taighde innealtóireachta, chun samhlacha urlabhra a chruthú (Caibidil 7) lenár féidir feasacht na teanga labhartha a chur i bhfeidhm sa teicneolaíocht, agus a aistriú go díreach d'áiseanna, m.sh do theagasc na teanga.

.



Figiúr 2.1: 2 phróiseas i nginiúint na cainte,
(a) giniúint na fuaime (an glór-fhoinse),
(b) scagadh na fuaime (an urlabhraíocht)

2.4.2 Urlabhraíocht Consan agus Gutaí na Gaeilge (Scagadh na Fuaime)

Is mó anailís cluaise atá déanta thar na blianta ar chonsain agus ar ghutaí na Gaeilge, agus tá foinsí saibhre eolais againn ar chóras fuaimeanna na gcanúintí a théann i bhfad siar [4 - 12]. Tá córas fairsing consan sa Ghaeilge, le hidirdhealú idir dhá shraith consan – caol agus leathan, m.sh. na consain / $l^{ɣ}$ / agus / l^{j} / i dtús na bhfocal / $l^{ɣ}$ o: $n^{ɣ}$/ lón agus / l^{j} o: $n^{ɣ}$/ leon (tuilleadh eolais in [2]). Ní hionann ceachtar acu agus an fhuaim /l/ i /lo:n/ 'lone' an Bhéarla. Bíonn an-tionchar ag an chonsan ar an ghuta atá taobh leis. Tá impleachtaí ag líon mór na gconsan Gaeilge do dhearadh agus do mhéid na gcorpas urlabhra a mbeidh gá leo don teicneolaíocht: mar gheall ar an tionchar ag fuaim amháin ar a chomharsa, is gá taifead a dhéanamh ar gach uile fhuaim, i ngach comhthéacs ina bhfaightear í (féach Caibidil 3 agus 8).

Má dhéantar malartú ó chonsan caol go leathan, athraíonn brí an fhocail, mar a fheictear sa sampla thuas. In amanna is brí ghramadúil an fhocail a athraíonn, m.sh le malartú / $n^{ɣ}$ / go / n^{j} / i ndeireadh an fhocail leon (/ l^{j} o: $n^{ɣ}$/ leon > / l^{j} o: n^{j}/ leoin) aistríonn sé ó uatha go hiolra, nó aistríonn an tuiseal ó ainmneach go ginideach (féach [2] agus §2.5.1 thíos).

Is beag anailís eimpíreach atá déanta go fóill le modhanna anailíse nuálacha ar na tréithe acústaice agus urlabhraíochta a bhaineann le consain agus gutaí na Gaeilge. Tá tábhacht lena leithéid, go háirithe mar gheall ar na difríochtaí móra idir an Ghaeilge agus

mórtheangacha amhail an Béarla, ar a bhfuil formhór na teicneolaíochta bunaithe. Tá anailís déanta ar ghnéithe faoi leith d'urlabhraíocht agus de thréithe acústaice na gconsan in [13-19], agus tá spléachadh le fáil ar na difríochtaí idir na canúintí [20-22].

Don todhchaí agus don anailís urlabhraíochta/acústaice ar chóras fuaimeanna na Gaeilge atá in oiriúint don teicneolaíocht, moltar díriú ar mhodhanna anailíse úrscothacha a bhaineann úsáid as teicnící mar an íomháú athshondais mhaighnéadaigh (fMRI). As a leithéid, tiocfaidh tuiscint nua ar urlabhraíocht na Gaeilge, ar na difríochtaí idir canúintí, agus leagfar dúshraith do bhunteicneolaíocht urlabhra agus d'áiseanna éifeachtacha do theagasc agus do theiripe na cainte. Mar atá luaite, beidh dlúthchomhoibriú idir teangeolaithe agus innealtóirí de dhíth don obair seo (féach Caibidil 7)

2.4.3 Prosóid na Gaeilge: an Glór-Fhoinse

Baineann an phrosóid le síormhodhnú an ghlóir, lena gcruthaítear patrúin sna foinn, i gcáilíochtaí an ghutha, in aiceann na siollaí agus i rithim na cainte. Mar a míníodh thuas, baineann muid saghasanna éagsúla eolais as na patrúin ghutha seo. Ar an chéad dul síos, ceadaíonn siad don éisteoir díchódú a dhéanamh, i.e., na focail, na frásaí, na habairtí a aimsiú i sruth na cainte. Tugann na patrúin leid dúinn ar na focail is tábhachtaí san abairt agus eolas eile ar an chiall atá le baint as abairt. Níl nótáil ar mhóran den eolas seo sa leagan scríofa. Ar an dara dul síos, mar atá mínithe cheana, léiríonn an síormhodhnú ar cháilíochtaí an ghutha gnéithe tábhachtacha de bhrí na cainte, i.e. mothúcháin agus meon an chainteora. Chomh maith leis sin, rialaíonn an phrosóid an comhrá idir beirt, ina dtugann an cainteoir nod don éisteoir cár cheart di/dó labhairt. Is gnéithe lárnacha den chumarsáid daonna iad seo, ach níl aon nótáil orthu sa scríobh.

Go dtí le déanaí, is beag eolas a bhí ar fáil ar phrosóid na Gaeilge, i gcomparáid leis na consain agus na gutaí. Tá anailís cluaise ar fhoinn (tuiníocht) chanúint Chonamara ar fáil in [23, 24], agus bhí plé an-ghairid ar phointí tuiníochta sna cuntais foghraíochta a tugadh ar chanúintí áirithe [6,7]. Le roinnt blianta anuas, tá taighde eimpíreach á dhéanamh i gColáiste na Tríonóide ar thuiníocht agus aiceann na gcanúintí . Is léir go bhuil difríochtaí móra eatarthu: i dtuaisceart na tíre (i nDún na nGall) téann foinn na tuiníochta in airde le béim na siollaí; i gConamara agus i gCiarraí, titeann siad go hiondúil [25-30]. Fiú i gcás canúintí gaolmhara, mar Inis Mór agus Cois Fharraige, is cosúil go bhfuil difríochtaí suntasacha ó thaobh tuiníochta de [31]. Is léir go bhfuil gá le taighde a chlúdaíonn an phrosóid i ngach uile chanúint.

Is ar na foinn agus ar aiceann na cainte a díríodh go dtí seo agus ar an ról atá acu i ndíchódú na cainte. Don todhchaí, beidh anailís i bhfad níos leithne de dhíth, ina bhfuil clúdach ar cháilíochtaí an ghutha agus ar na gnéithe prosóide a bhaineann le léiriú na mothúchán agus stiúradh na cumarsáide idirphearsanta. Beidh sé seo tábhachtach do theicneolaíocht na hurlabhra sa todhchaí, ina mbeidh muid ag comhrá le carachtair fhíorúla, le róbait srl. (Caibidlí 8-10). Chun gur féidir anailís chruinn a dhéanamh ar cháilíochtaí an ghlóir, beidh modhanna anailíse de dhíth a bhaineann úsáid as samhlacha innealtóireachta (tá plé air seo i gCaibidil 7). Mar an gcéanna le taighde na hurlabhraíochta, moltar dlúthchomhoibriú taighde idir lucht foghraíocht-teangeolaíochta agus innealtóireachta chun eolas níos iomláine a fháil ar phrosóid na Gaeilge, agus chun a chintiú go leabófar an fheasacht seo i dteicneolaíochtaí agus in áiseanna urlabhra do phobal na Gaeilge.

2.4.4 Ón fhuaim go litriú

Sa scríobh seasann litreacha na haibítre do chonsain agus do ghutaí na teanga labhartha, agus is beag nótáil ar an phrosóid. Tá an chomhfhreagairt idir fuaimeanna agus litiriú na Gaeilge casta agus doiléir. Tá sé casta mar gur minic gur féidir fuaim áirithe a scríobh ar a lán bealaí difriúla. Tá sé doiléir sa mhéid is nach nochtar an difríocht idir na consain caola agus leathana ar bhealach díreach sa scríobh. (Is fiú a lua mar sin féin, go bhfuil an chomhfhreagracht idir fuaim-agus-litir i bhfad níos rialta ná mar atá sa Bhéarla). Tá impleachtaí ag na tréithe seo de chóras scríofa na Gaeilge do shealbhú na litearthachta agus pléitear an cheist seo in §2.7.3).

Faoi láthair, baintear úsáid as inchur/aschur téacs de ghnáth i gcórais sintéise agus aitheanta cainte. Mar sin, tá gá le rialacha a aistríonn ó litir-go-fuaim. Tá difríochtaí suntasacha sna rialacha seo do na canúintí éagsúla, agus is gá mar sin iad a chur in oiriúint do gach canúint. Tá forbairt déanta ar rialacha litir-go-fuaim do na trí mhórchanúint mar chuid d'fhorbairt ar ghlórtha sintéiseacha ABAIR i gColáiste na Tríonóide, agus tá an obair seo ag tacú le taighde ar shealbhú na litearachta [32] (féach §2.7.3 agus Caibidil 15).

2.5 Brí a chur in iúl trí struchtúr gramadaí

Is iad an mhoirfeolaíocht agus an chomhréir réimsí na hanailíse teanga a bhaineann le struchtúr agus brí focal, frásaí agus abairtí – réimsí a áirítear i gcuntais ghramadaí thraidisiúnta, agus ina mbíonn an anailís bunaithe de ghnáth ar fhoirmeacha scríofa.

2.5.1 Struchtúr na bhfocal: moirfeolaíocht

Maidir le teicneolaíocht teanga, tá tábhacht bhunúsach ag baint le bheith in ann focail (agus focail fhéideartha) na teanga a aithint. Is gá freisin a bheith in ann gnéithe focail a aithint, mar shampla a fhréamhfhoirm (ceannfhocal nó leama), agus cén cineál focail é, .i. cibé an briathar nó ainmfhocal é srl (ranna cainte). I gcás briathra tá sé tábhachtach an aimsir a aithint (aimsir chaite, aimsir láithreach, aimsir fháistineach srl), agus teastaíonn gnéithe eile amhail uimhir, pearsa, gné, modh srl freisin. I gcás ainmfhocal, tá sé tábhachtach a bheith in ann a aithint an bhfuil an t-ainmfhocal uatha nó iolra, agus tá tábhacht freisin le gnéithe eile amhail inscne (baininscneach nó firinscneach) agus tuiseal gramadaí (ainmneach, ginideach, gairmeach srl). Tá an cineál seo faisnéise riachtanach le haghaidh go leor feidhmchlár teanga coitianta, amhail litreoirí, seiceálaithe gramadaí, foclóirí, teasárais, (Caibidil 13) chomh maith le hinnill chuardaigh (Caibidil 11), feidhmchláir aistriúcháin (Caibidil 11), feidhmchláir oideachais (Caibidil 14), agus giniúint teanga (Caibidil 6).

Is éard is moirfeolaíocht teanga ann ná staidéar ar struchtúr inmheánach na bhfocal. Tá a córas rialacha féin ag gach teanga chun focail nua a thabhairt isteach sa teanga, agus chun na foirmeacha éagsúla focal a chruthú. De ghnáth, is éard a bhíonn i gceist leis an gcóras ná táthmhíreanna (iarmhíreanna agus réimíreanna, mar shampla) a chur leis an fhréamh nó an fhréamh a athrú nó an dá rud. Dá bhrí sin, tá go leor focal comhdhéanta d'fho-mhíreanna inaitheanta níos lú, i.e. an fhréamh agus na réimíreanna agus/nó na hiarmhíreanna. Tugtar 'moirféimí' ar na fo-ranna seo. Mar shampla, tá dhá mhoirféim ag an bhfocal tógfaidh: an fhréamh léacsach tóg agus an iarmhír aimsire fáistiní-faidh. Is féidir leis an fhréamh féin athrú freisin. I measc na n-athruithe fréamhacha tá athruithe tosaigh amhail séimhiú m.sh. chun an aimsir chaite a chur in iúl; thóg vs. tóg, agus mionathruithe inmheánacha amhail caolú (an t-athrú ó chonsan deiridh leathan go caol, m.sh. chun an fhoirm iolra a chur in iúl; báid vs. bád' (féach §2.4.2).

Toisc go n-athraíonn cruth bunfhréimhe focail (leama) go hiondúil, agus go háirithe toisc gur féidir tús an fhocail athrú, tá tábhacht ar leith do phróiseáil urlabhra agus teanga na Gaeilge ag baint le huirlisí próiseála teanga nádúrtha (PTN) amhail leimeoirí a aithníonn an fhréamh (i.e. an leama), agus clibeálaithe roinne cainte, a aithníonn an chatagóir roinne cainte agus na gnéithe a bhaineann léi.

Tá feiniméin mhoirfeolaíocht na Gaeilge doiciméadaithe go maith [33, 34] agus forbraíodh uirlisí moirfeolaíochta riailbhunaithe agus clibeálaithe roinne cainte (Caibidil 5.4) ar féidir leo leama agus gnéithe d'fhocal a aithint/a ghiniúint agus clibeanna roinne cainte a shannadh d'abairt nó do théacs [35]. Is iad na dúshláin is mó atá roimh anailís mhoirfeolaíoch na Gaeilge i gcomhthéacs na teicneolaíochta teanga ná an córas a choinneáil suas chun dáta i gcónaí de réir mar a thagann focail nua isteach sa teanga. Is féidir é sin a bhaint amach trí naisc láidre a bhunú le bunachair léacsacha (Caibidil 4). D'fhéadfaí feabhas a chur ar na huirlisí mhoirfeolaíochta riailbhunaithe trí níos mó míreanna ilfhoclacha a ionchorprú, eintitis ainmnithe amhail ainmneacha daoine, áiteanna agus eagraíochtaí (Caibidil 5.5) agus comhtharluithe coitianta focal (Caibidil 3.2).

Is féidir modhanna staitistiúla (seachas modhanna riailbhunaithe), corpasbhunaithe a úsáid i gcásanna nuair nach gcloíonn teanga i gcónaí leis na rialacha moirfeolaíochta caighdeánacha, m.sh. sonraí meán sóisialta. Maidir le próiseáil teanga staitistiúil agus teicneolaíocht meaisínfhoghlama, tá dúshlán ag baint le córas moirfeolaíoch saibhir na Gaeilge, áit a bhféadfadh go leor foirmeacha labhartha agus scríofa a bheith ag focal amháin ag brath ar an gcomhthéacs gramadaí. Úsáidtear an téarma 'ganntanas sonraí' chun cur síos a dhéanamh ar an nasc idir an mhoirfeolaíocht agus an tionchar a bhíonn aici ar mhéideanna riachtanacha corpas i modhanna atá bunaithe ar staitisticí (féach Caibidil 2.7). Ciallaíonn sé seo nuair atá corpas á thógáil (Caibidil 3) le haghaidh na foghlama staitistiúla nó le haghaidh meaisínfhoghlama, mar shampla meaisínaistriúchán staitistiúil (Caibidil 11) nó aithint cainte (Caibidil 9), nach mór do chorpais a bheith an-mhór d'fhonn clúdach leordhóthanach a chur ar fáil ar fhoirmeacha uile focal na teanga.

2.5.2 Struchtúr na n-abairtí: comhréir

Go minic, ní leor fios a bheith againn faoi airíonna na bhfocal in abairt, ní mór dúinn a fháil amach freisin cad é an ról atá acu in abairt. Tá sé seo tábhachtach don tuiscint bhunúsach ar 'cé a rinne cén rud ar cén duine'. Baineann comhréir le struchtúr abairtí, agus, go háirithe, le córas na rialacha a chinneann cén dóigh a gc uitear focail le chéile chun frásaí a chruthú agus cén dóigh a gcuirtear frásaí le chéile chun abairtí a chruthú. Baineann sé chomh maith le feidhmeanna gramadaí (Caibidil 3.4.1, corpas scríofa) amhail ainmní agus cuspóir, a shainaithint, agus chun na gaolta ordlathacha idir frásaí in abairt a chinneadh. Tugtar parsáil ar an gcineál seo anailíse, agus tugtar parsálaí ar ríomhchlár a dhéanann anailís chomhréireach uathoibríoch ar theanga. Tá parsáil ríthábhachtach in go leor feidhmchlár próiseála teanga nádúrtha, amhail seiceáil gramadaí,

feidhmchláir FRT, ríomhaistriúchán riailbhunaithe, agus tuiscint agus giniúint teanga nádúrtha. Tóganna an pharsáil ar anailís mhoirfeolaíoch agus clibeáil roinne cainte a luaitear sa mhír roimhe seo.

(1) Léigh an cailín an litir
Read the girl the letter
'The girl read the letter'

Is minic a aithnítear ainmní agus cuspóir gramadaí abairte i nGaeilge (agus i mBéarla) ón áit a nglacann sé san abairt. Mar shampla, in abairt (1), baintear úsáid as an mbriathar 'léigh' agus as dhá fhrása ainmfhoclacha ina dhiaidh; 'an cailín' , agus 'an litir' . Sa Ghaeilge, tá a fhios againn gurb é an chéad fhrása ainmfhoclach i ndiaidh an bhriathair seo an t-ainmní gramadúil agus gurb é an dara frása ainmfhoclach an cuspóir gramadúil.

(2) Léigh an cailín clúdach na litr-each
Read the girl cover the letter-GENITIVE
'The girl read the envelope' i.e. the cover of the letter

In abairt (2), tá a fhios againn gur aonad amháin é clúdach na litreach, i.e. frása ainmfhoclach amháin a bhfuil feidhm ghramadaí 'cuspóra' aige de bharr an ghaoil tuisil ghinidigh idir na hainmfhocail clúdach agus litreach.

Sa phróiseáil teanga nádúrtha, tá tábhacht ag baint le bheith in ann caidrimh chomhréireacha a aithint chun brí abairte a thuiscint. Tá sé seo riachtanach le haghaidh aisghabháil faisnéise agus achoimriú doiciméad (Caibidil 12), agus tá sé thar a bheith tábhachtach le haghaidh meaisínaistriúchán ar phéirí teanga amhail Béarla agus Gaeilge, a bhfuil ord éagsúil focal acu. Déantar cur síos ar uirlisí parsála agus smutála don Ghaeilge i gCaibidil 5.

Maidir le comhréir na Gaeilge, is aisteach nach bhfuil mórán sonraí le fáil sna saothair thagartha ghramadaí thraidisiúnta [33, 36]. Rinneadh formhór na staidéar teoiriciúil ar chomhréir na Gaeilge laistigh de chreat giniúnach Chomsky [37-41] agus laistigh den chreat fheidhmiúil Gramadaí Róil agus Tagartha (GRT) [42]. Níor ceapadh na creataí seo chun críocha ríomhaireachta, dá bhrí sin le fiche nó tríocha bliain anuas forbraíodh roinnt samhlacha teoiriciúla gramadaí agus comhréire atá dírithe ar an ríomhaireacht, go háirithe Gramadach Fheidhmiúil Léacsach (GFL) [43] agus Gramadach Struchtúir Frása Ceann-Thiomáinte (GSFC) [44]. Aithníonn na creataí comhréireacha seo an gá atá le leibhéil éagsúla anailíse teanga a chomhtháthú, lena n-áirítear moirfeolaíocht, séimeantaic agus pragmataic agus an gá lena gcuid anailísí a chur i láthair ar bhealach ríomhaireachtúil. Tá aghaidh tugtha ar ghnéithe áirithe de chomhréir na Gaeilge trí úsáid a bhaint as an gcreat GFL [45, 46]. Le blianta beaga anuas, tá an-tóir ar chomhréir spleáchais [47] chomh maith mar chreat ríomhaireachtúil comhréire.

Cé go ndearnadh go leor taighde ar chomhréir theoiriciúil an Bhéarla agus mórtheangacha eile na hEorpa, ní gá go n-aistreodh sé seo go comhréir na Gaeilge mar gheall ar na difríochtaí suntasacha idir na teangacha ó thaobh ord na bhfocal agus nathanna cainte de. Tá sé riachtanach tuilleadh taighde a dhéanamh ar chomhréir na Gaeilge, go háirithe laistigh de creat ríomhaireachtúil, chun bonn láidir a chur faoi pharsáil ríomhaireachtúil, faoi ghraiméir thagartha don teanga agus faoi fheidhmeanna FRT srl.

2.6 Brí dhúchasach agus iomlánaíoch

Baineann staidéar na séimeantaice le brí dhúchasach na bhfocal agus a ngaol lena gcéile (m.sh. an fhaisnéis a stóráiltear i mbunachair eolais léacsacha mar a shonraítear i gCaibidil 4), agus baineann an phragmataic leis na dóigheanna a dtéann an comhthéacs i bhfeidhm ar an bhrí. Go praiticiúil tá dearcadh níos iomlánaíche againn ar bhrí, mar go dtuigeann muid focail agus abairtí ar dhóigheanna difriúla i gcomhthéacsanna difriúla. Sa chomhrá labhartha, tugann an phrosóid agus tuin ghutha an chainteora go háirithe cuid mhór den bhrí iomlán.

2.6.1 Brí dhúchasach: séimeantaic

Ionas go mbeidh córas ríomhaireachtúil in ann abairt a thuiscint, i.e. struchtúr áirithe a shannadh d'abairt, is gá faisnéis shéimeantach a bheith ann a bhaineann le brí focal. Mar shampla, i gcás briathar áirithe tá sé úsáideach fios a bheith againn cé mhéad rannpháirtí a bhíonn páirteach sa ghníomh de ghnáth, i.e. struchtúr faisnéis-argóinte an bhriathair. Cabhróidh sé seo a bhunú cé mhéad frása ainmfhocal is féidir linn a bheith ag súil leis san abairt. Tá sé úsáideach freisin go mbeadh a fhios againn cad iad na róil a chomhlíonann na rannpháirtithe seo, mar go mbainfidh sé seo le feidhmeanna gramadúla an ainmní agus an chuspóra srl. Tá faisnéis faoi bhriathra i dtéarmaí líon na rannpháirtithe a theastaíonn uathu, agus na róil shéimeantacha a theastaíonn, ionchódaithe i bhfoclóir fiúsachta. Tugtar srianta nó roghanna roghnaitheacha na faisnéise ar fhaisnéis faoi na cineálacha ainmfhocal ar féidir leo na róil shéimeantacha sin a chomhlíonadh. Mar shampla, tá cuspóir díreach de dhíth ar bhriathar amhail 'ól', ach is féidir linn a rá freisin gur cheart gur leacht é an cuspóir sin. Chomh maith le líon na n-eintiteas/rannpháirtithe a theastaíonn, d'fhéadfadh an tuiscint a bhaintear as abairt a bheith ag brath ar shaintréithe na n-eintiteas féin.

(3) Tá sí ina gairdín
Is she in.her garden
'She is in her garden'

(4) Tá sí ina dochtúir

Is she in.her doctor
'She is a doctor'

Mar shampla, tá struchtúr argóinte comhchosúla na ag abairtí (3) agus (4), ach baineann muid ciall astu ar dhóigh dhifriúil bunaithe ar shaintréithe na n-ainmfhocal i gceist. Tuigtear 'gairdín' mar shuíomh agus tuigtear 'dochtúir' mar ghairm, i.e. bíonn airíonna séimeantacha difriúla ag na hainmfhocail. Dá bhrí sin, is gá faisnéis a dhoiciméadú freisin faoi chatagóirí ainmfhocal. Nuair a bhíonn a fhios againn an seasann ainmfhocal do dhuine, do ní beo nó neamhbheo, leacht, soladach, suíomh srl., éascaíonn sé parsáil láidir shioncronach agus tuiscint shéimeantach ar bhrí. Ina theannta sin, is féidir caidrimh shéimeantacha chórasacha amhail comhchiallaigh (focail a bhfuil an bhrí chéanna leo), focail frithchiallaigh (focail a bhfuil a mhalairt de bhrí leo) agus comhainmneacha (aicmí agus fo-aicmí) a ionchódú i mbunachair eolais (ar a dtugtar 'gréasáin focal' freisin), chun faisnéis shéimeantach luachmhar a sholáthar d'iarratais PTN (féach Caibidil 4.3).

Is beag taighde nó cáipéisíocht theangeolaíoch atá déanta go dtí seo ar shéimeantaic na Gaeilge. Ní mór an cineál taighde séimeantach atá ag teastáil le haghaidh na teicneolaíochta urlabhra agus teanga a dhéanamh i gcomhthéacs creataí ríomhaireachtúla oiriúnacha. Ní mór di a bheith bunaithe ar chorpas (Caibidil 3) agus beidh sainuirlisí próiseála teanga ag teastáil lena haghaidh (Caibidil 5). Sna caibidlí a leanann an chaibidil seo, déanaimid cur síos ar chorpas nótáilte séimeantach (Caibidil 3), ar fhoclóirí fiúsanna agus ar líonraí séimeantacha (Caibidil 4), agus ar uirlisí anótála ar nós lipéadóirí róil shéimeantaigh agus aitheantóirí ainmneacha eintitis (Caibidil 5).

2.6.2 Brí iomlánaíoch: pragmataic

Ní hionann brí abairte nó cainte agus suim na bhfocal aonair le chéile, an struchtúir shintéisigh nó an bhrí shéimeantach bhunúsach. Braitheann an bhrí freisin ar an gcomhthéacs ina n-úsáidtear é, agus i gcás na teanga labhartha, ar an bprosóid agus ar ghuth an chainteora. Breathnaíonn pragmataic ar ghnéithe níos iomlánaíche na brí, agus oibríonn sí ar leibhéal an "dioscúrsa ", áit a dtagann abairtí nó cora cainte le chéile chun téacs nó comhrá comhleanúnach a chruthú. Is fiú machnamh a dhéanamh i staidéar na pragmaitice ar an gcaidreamh idir an cainteoir agus an t-éisteoir, nó idir an scríbhneoir agus an léitheoir, mar go mbeidh tionchar ag an gcaidreamh ar an méid foirmeáltachta atá oiriúnach. Beidh an bhrí i gcomhthéacs ar leith ag brath freisin ar réamheolas agus toimhdí na rannpháirtithe, chomh maith le "heolas ar an domhan" i gcoitinne. Díríodh go leor taighde idirnáisiúnta ar bhealaí iomchuí agus ríomh-éifeachtúla a aimsiú chun eolas ar an domhan a ionchódú chun cabhrú le tuiscint teanga nádúrtha, ach is beag taighde atá déanta ar phragmataic na Gaeilge go dtí seo.

I ndioscúrsa scríofa agus labhartha, seachnaítear athrá go ginearálta chun níos lú focal a úsáid agus chun aird a tharraingt ar an difríocht idir faisnéis nua agus faisnéis a luadh roimhe seo. Is minic a chiallaíonn sé seo go n-úsáidtear forainmneacha in áit na n-ainmfhocal thuasluaite. Ceann de na dúshláin a bhaineann le feidhmchláir um thuiscint teanga nádúrtha ná an chiall cheart a bhaint as na forainmneacha, m.sh. fáil amach cén frása ainmfhoclach (nó ráiteas) dá dtagraíonn forainm. Bíonn an freagra le fáil de ghnáth sa téacs (dioscúrsa) nó i réimse an eolais ghinearálta. Mar shampla, i gcás Chonaic mé é, tagraíonn an forainm é do dhuine nó rud éigin a luadh roimhe sin, i.e. d'fhéadfaí an abairt seo a léiriú mar 'I saw it' nó 'I saw him'. Cinnfidh an dioscúrsa a thagann roimh an abairt seo, mar aon le heolas ar an domhan agus na toimhdí atá i bpáirt againn lena chéile, an tagairt is dóichí don fhorainm sa chomhthéacs áirithe seo. Mar an gcéanna, in abairt mar 'Thug an t-uachtarán óráid', d'fhéadfadh 'an t-uachtarán' tagairt a dhéanamh d'uachtarán tíre, eagraíochta nó gnólachta, agus braitheann sé ar chúrsaí ama freisin. Tugtar 'réiteach iarthagartha' ar an bpróiseas trína ndéantar tagráin forainmneacha a chinneadh i dtéacs nó i dtras-scríbhinn. Is den riachtanas é eolas a fháil ar na prionsabail a bhaineann le struchtúr an dioscúrsa agus le pragmataic sa Ghaeilge chun forbairt a dhéanamh ar na cineálacha uirlisí a d'fhéadfadh an tasc PTN seo a chomhlíonadh (Caibidil 5). Bíonn tábhacht ag baint leis an réiteach iarthagartha don Tuiscint Teanga Nádúrtha, go háirithe i bhfeidhmeanna oideachais (Caibidil 14) agus don aisghabháil faisnéise (Caibidil 12)

Sa chaint, is iompróir príomhúil faisnéise pragmataí í an prosóid ghutha (§2.4.3 thuas agus Caibidil 7). Mar shampla, cuireann fócas prosóideach in iúl cé acu cuid den chaint atá tábhachtach agus cén chuid nach bhfuil (i.e. eolas roimh ré). Is tríd an nglór a chuirtear in iúl freisin cén chiall atá le baint as an méid a dúradh, m.sh. mar ráiteas, mar cheist, mar ordú nó mar fhógairt. Rialaíonn comharthaí prosóideacha an comhrá, agus tugann siad le fios don éisteoir cathain atá a sheal cainte féin ann. Agus an chuid is bunúsaí, b'fhéidir, is ea go léiríonn an phrosóid (tuin an ghutha go háirithe) caidreamh an chainteora leis an éisteoir (m.sh. foirmiúil, pearsanta, agus mórchúiseach). I gcomhrá, tagann prosóid na gcainteoirí i ngar dá chéile agus téann siad ar shiúl óna chéile, rud a léiríonn éabhlóid an naisc atá eatarthu agus an chaoi a dtaispeánann siad a gcuid mothúchán (brónach, corraithe, faoi leadrán).

D'fhéadfadh focail agus tuin an ghutha nach dtagann le chéile a bheith ina chomhartha searbhas. Tá na gnéithe seo den bhrí riachtanach le haghaidh cumarsáid éifeachtach labhartha, agus le haghaidh feidhmeanna teicneolaíochta urlabhra saibhrithe (féach Caibidlí 7, 8, 10).

I réimse an teagaisc agus measúnaithe teanga (§2.7.2), is léiriú é ardchumas pragmataice ar ardchumas teanga. Dá bhrí sin, tá sé tábhachtach go mbeifí in ann tógáin phragmataice a aithint, mar shampla tá tábhachtach ag baint le húsáid chuí a bhaint as nathanna cainte, iarratais indíreacha, agus gnéithe dioscúrsa cuí eile don teagasc agus measúnú teanga. Is féidir corpais scríofa agus labhartha anótáilte (Caibidil 3.3) a anailísiú d'fhonn tógáin phragmataice a aithint. Ina theannta sin, tá corpais anótáilte foghlaimeoirí (Caibidil 3.5) riachtanach chun forbairt inniúlachta pragmataice i bhfoghlaimeoirí a rianú le himeacht ama. Tá sealbhú na prosóide amhail na prosóide dúchasaí riachtanach dá n-inniúlacht phragmataice sa teanga labhartha. Teastaíonn taighde trasdisciplíneach comhordaithe chun a chinntiú gur féidir eolas foghraíochta (§2.4.3) a aistriú go foirm atá oiriúnach don teagasc, agus a leabú i bhfeidhmeanna oideolaíocha (Caibidil 7, 14).

Ní mór d'fheidhmchláir theicneolaíochta urlabhra agus teanga ar gá dóibh teanga a ghiniúint, mar théacs nó mar chaint, teanga a tháirgeadh atá cruinn ó thaobh na gramadaí de agus a oireann ó thaobh na pragmataice de. Tá tuiscint agus giniúint cheart an dioscúrsa an-tábhachtach do chórais idirphlé labhartha agus scríofa (Caibidil 10), ríomhaistriúchán (Caibidil 11), chomh maith le córais freagartha ceisteanna (Caibidil 12). Tá prosóid ghutha (§2.4.3) ríthábhachtach maidir le córais idirghníomhacha idirphlé labhartha, atá ag éirí níos gaire ná riamh d'idirghníomhaíocht labhartha an duine (Caibidlí 7, 8, 10). Dá bhrí sin, is cuid riachtanach de na feidhmeanna urlabhra agus teicneolaíochta teanga níos sofaisticiúla í an phragmataic.

Go dtí seo is beag taighde atá déanta maidir leis na gnéithe pragmataice d'úsáid na Gaeilge, sa teanga scríofa nó labhartha. Ní mór do staidéir phragmataice théacsbhunaithe chumarsáide díriú ar na gnéithe pragmataice den dioscúrsa scríofa agus ar an dioscúrsa labhartha tras-scríofa. Chun go mbeidh an teicneolaíocht urlabhra in ann teacht ar an mbrí iomlánaíoch in idirghníomhaíochtaí labhartha, teastaíonn taighde ar an ngné phragmataice den ghuth-phrosóid, rud a chuirfeadh leis na hanailísí téacsbhunaithe thuasluaite. Tá gá le taighde bunúsach, toisc go bhfuil feasacht ar an bpragmataic tábhachtach do réimse leathan teicneolaíochtaí urlabhra agus teanga. Is féidir le taighde ar anphragmataic tacú le dearadh curaclaim, teagasc agus measúnú don Ghaeilge freisin.

2.7 Leibhéil Sealbhaithe agus Inniúlachta Teanga agus Litearthachta

Tá taighde ar shealbhú na Gaeilge ina bhunús riachtanach do theicneolaíochtaí oideachais a bhfuil an teangeolaíocht mar bhonn eolais fúthu. Chomh maith leis an teicneolaíocht teanga, beidh go leor buntáistí tábhachtacha eile ag an taighde seo do mhúineadh na Gaeilge agus do dhiagnóisiú laguithe teanga i gcomhthéacs shealbhú na Gaeilge mar chéad teanga.

Tá an próiseas trína bhfoghlaimítear an chéad teanga sa luath-óige, agus an bealach ina bhfoghlaimítear an dara teanga i suíomh oideachais, difriúil ar bhealaí éagsúla. Mar thoradh air seo, tá réimsí ar Leith staidéir ann maidir le sealbhú na chéad teanga (fo-réimsí maidir le sealbhú aonteangach, dátheangach agus ilteangach) agus maidir le sealbhú na dara teanga.

2.7.1 2.7.1 Sealbhú Céadteanga & Sealbhú Céadteanga Dhátheangach

Ceann de na bealaí ina ndéantar staidéar ar shealbhú céadteanga ná trí shamplaí de theanga labhartha spontáineacha a bhailiú le haghaidh anailíse. De ghnáth, déantar taifid ar idirghníomhaíochtaí idir leanbh agus an cúramóir go tráthrialta i suíomh nádúrtha, le linn na chéad chúpla bliain de shaol an linbh. Rinneadh roinnt staidéir ar shealbhú teanga na Gaeilge mar chéad teanga i bpáistí le himeacht na mblianta [48 -51] agus bailíodh sonraí luachmhara. Mar sin féin, i gcomparáid le teangacha eile ar nós an Bhéarla, is beag atá ar eolas faoi cathain agus cén t-ord ina bhfaigheann an cainteoir dúchais gnéithe éagsúla de chórais fuaime, moirfeolaíochta agus gramadaí na Gaeilge. Is rómhinic a ghlactar leis go mbeidh an próiseas sealbhaithe Gaeilge (mar chéad nó mar dhara teanga) cosúil leis an bpróiseas sealbhaithe Béarla nó teangacha eile a bhfuil staidéar maith déanta orthu. Níl call leis an toimhde seo ar chor ar bith, i bhfianaise na ndifríochtaí móra struchtúracha idir teangacha.

Tá gá le hanailísí eimpíreacha spriocdhírithe ar shonraí sealbhaithe céadteanga chun aird a tharraingt ar an ghnáth-dhul chun cinn i dtaobh shealbhú na Gaeilge mar chéad teanga i dtimpeallacht dhátheangach/aonteangach agus ar na clochmhílte ábhartha. Is éard a bheidh i gceist leis seo ná sonraí breise a bhailiú, le cur leis na sonraí atá ann cheana. Is bunchloch riachtanach é an buntaighde seo do chórais FTR éifeachtach agus is féidir leis an taighde seo a bheith mar bhonn agus mar thaca ag dearadh curaclaim teanga do scoileanna Gaeilge i scoileanna Gaeltachta agus i nGaelscoileanna araon.

2.7.2 Sealbhú na Dara Teanga agus Leibhéil Inniúlachta Teanga

Is cinnte go bhfuil tionchar ag an gcéad teanga ar eispéireas an fhoghlaimeora agus iad i mbun teanga nua a fhoghlaim.

Baineann an deacracht a bhíonn ag an bhfoghlaimeoir an dara teanga a shealbhú le cé chomh difriúil agus atá an dara teanga óna gcéad teanga. De ghnáth, is iad na gnéithe den dara teanga nach bhfuil i láthair sa chéad teanga na deacrachtaí is mó don fhoghlaimeoir. Tá difríocht shuntasach idir an Ghaeilge agus an Béarla, mar atá luaite cheana. Ceann de na bealaí chun na dúshláin a chruthaíonn sé seo a shárú ná anailís mhionsonraithe a dhéanamh ar dhul chun cinn na bhfoghlaimeoirí dara teanga, trí chorpais foghlaimeoirí a bhailiú agus a anailísiú (Ch. 3.5) ag na leibhéil éagsúla oideachais.

Ós rud é go sealbhaíonn formhór na ndaoine an Ghaeilge mar dhara teanga, go hiondúil tar éis dóibh an Béarla a shealbhú go haonteangach, tá géarghá le taighde sa réimse seo chun bonn eolais a chur faoi fhorbairt ar theicneolaíochtaí oideachais cuí. De bhreis ar chórais teicneolaíochta oideachais tógtha ar bhonn láidir, tacóidh an taighde seo freisin le dearadh curaclaim do mhúineadh agus measúnú na Gaeilge i scoileanna príomhshrutha.

Ábhar iontais a laghad taighde atá déanta ar shealbhú na Gaeilge mar dhara teanga (nó teanga ina dhiaidh sin) i suíomh scoile nó i measc foghlaimeoirí fásta. Ceann de na bealaí ina ndéantar staidéar ar shealbhú na dara teanga ná trí anailís a dhéanamh ar chorpais foghlaimeoirí teanga [52] (Ch. 3.5), agus beidh gá le corpas breise foghlaimeora i suíomhanna éagsúla. Tá go leor taighde ar siúl faoi láthair ar shealbhú na dara teanga i gcomhthéacs Chomhchreata Tagartha na hEorpa um Theangacha (CEFR) [53-55]. Úsáideann Comhchreat Tagartha na hEorpa um Theangacha leibhéil A1 agus A2 (bunúsáideoirí), B1 agus B2 (úsáideoirí neamhspleácha) agus C1 agus C2 (úsáideoirí inniúla) agus aithníonn sé na scileanna ar leith a bhaineann le labhairt, éisteacht, idirghníomhú, léamh agus scríobh. Is gnách le foghlaimeoirí teanga ag pointe áirithe leibhéil éagsúla inniúlachta a bheith acu sna scileanna éagsúla seo.

Sonraítear leibhéil inniúlachta A1 go C2 de chuid Chomhchreata Tagartha na hEorpa um Theangacha faoin gcreat ar bhealach neamhspleách ó thaobh teanga de ag baint úsáid as ráitis 'Can Do' . Ciallaíonn sé sin nach mór na ráitis 'Can do' do gach leibhéal de Chomhchreat Tagartha na hEorpa um Theangacha a mhapáil chuig eilimintí ar leith den fhóineolaíocht, moirfeolaíocht, comhréir, foclóireacht agus dioscúrsa srl. sa teanga i gcás teanga ar leith, cuir i gcás an Béarla nó an Ghaeilge. Is féidir mapáil a dhéanamh ar ráitis 'Can Do' le gnéithe ar leith den teanga ar an bhealach is éifeachtúla trí úsáid a bhaint as acmhainní agus uirlisí digiteacha. Is féidir anailísí mionsonraithe corpais a úsáid chun leibhéil Comhchreata Tagartha na hEorpa um Theangacha don Ghaeilge a thagarmharcáil, trí mhíreanna sonracha foclóra, struchtúir ghramadaí agus gnéithe teanga labhartha a chinneadh, a d'fhéadfadh a bheith bainteach le gach leibhéal inniúlachta.

Tá taighde nuálach corpas-bhunaithe ar leibhéil Comhchreata Tagartha na hEorpa um Theangacha agus ar an Ghaeilge á dhéanamh i gColáiste na Tríonóide, Baile Átha Cliath agus in Ollscoil Mhá Nuad, ag baint úsáide as corpas ábhair teagaisc [56], corpas foghlaimeoirí scríofa agus labhartha [52], mar aon le corpas tagartha ginearálta (Caibidil 3).

Chomh maith leis an tábhacht a bhaineann le tagarmharcáil Comhchreata Tagartha na hEorpa um Theangacha chun feidhmchláir chuí oideolaíocha FRT a fhorbairt, tá an taighde seo ríthábhachtach chun curaclaim agus measúnacht chomhtháite teanga a fhorbairt a thacóidh le teagasc agus foghlaim na Gaeilge ag gach leibhéal oideachais agus a chuirfidh feabhas orthu. I gCaibidil 3, tugtar mionchuntas ar na cineálacha corpais foghlaimeoirí agus cainteoirí líofa atá ag teastáil le haghaidh taighde ar shealbhú na Gaeilge sa chéad teanga agus sa dara teanga, mar aon leis na huirlisí anailíse agus anótála teanga (Caibidil 5) a bheidh ag teastáil chun tacú leis an taighde seo.

2.7.3 Sealbhú na Litearthachta

Tagann an chaint go nádúrtha don pháiste, ach caitear scileanna na léitheoireachta agus na scríbhneoireachta a theagasc. Do shealbhú na litearthachta tá tuiscint de dhíth ar (i) chóras fuaimeanna na teanga (gné lárnach den fheasacht fóineolaíochta) agus ar (ii) an nasc idir na litreacha scríofa agus fuaimeanna na teanga (feasacht fónaice). Ní hionann do gach teanga: nuair atá an chomhfhreagracht idir fuaim agus litriú soiléir agus rialta mar atá sa Spáinnis, foghlaimíonn an páiste léitheoireacht i bhfad níos tapúla ná mar a tharlaíonn sa chás ina bhfuil an chomhfhreagracht doiléir nó mírialta, mar atá sa Bhéarla [57].

I gcás na Gaeilge, tá (i) córas fuaimeanna fairsing, ina ndéantar idirdhealú idir consain chaola agus leathana, rud nach bhfuil sa Bhéarla (féach §2.4.2 agus [2]). Chomh maith leis seo (ii) tá an chomhfhreagracht idir litir agus fuaim (córas fónaice) doiléir agus casta go leor. Fáth amháin leis seo ná nach neosann an aibítir (Rómhánach) an difríocht idir na consain chaola/leathana ar na litreacha do na consain. Léirítear an difríocht idir consain trí rogha an ghuta a luíonn taobh leis: mar shampla, san fhocal 'buíon', /bɤiːnɤ/ léiríonn na litreacha "u" agus "o" gur consain leathana iad na consain ag tús agus ag deireadh an fhocail [2]. Fáth eile le doiléire an chórais scríofa ná go dtéann sé siar i bhfad: athraíonn fuaimniú na bhfocal thar am, ach ní athraíonn an leagan scríofa ag an luas céanna. Agus athraíonn an fuaimniú ar bhealaí difriúla i gcanúintí éagsúla: féach mar shampla mar a deirtear an focal (ní)

bhfaighfidh i gcanúintí difriúla: > /ʊɤaɪgʲ/ (i gCiarraí), /wɤaɪ/ (i gConamara), /wɤiː/ (i dTír Chonaill).

Ach cé go bhfuil an nasc idir litir agus fuaim casta, tá sé rialta ar an iomlán (agus i bhfad níos rialta ná córas fónaice an Bhéarla). Mar gheall air seo, is léir gur ar bhunús fonaice is fearr a d'fhéadfaí an léitheoireacht a theagasc. Faraor, ní mar seo atá: is meabhrú ar fhocail iomlána an modh a úsáidtear de ghnáth [58, 59]: tá seo mí-oiriúnach don Ghaeilge, go háirithe mar go bhfuil mórán foirmeacha ag focail sa Ghaeilge (féach §2.5.1).

Dóibh siúd atá ag foghlaim léamh agus scríobh an Bhéarla ag an am céanna (nó a bhfuil córas an Bhéarla foghlaimaithe acu cheana féin) tá sé tábhachtach go dtuigtear na difríochtaí idir an dá theanga, i dtaca leis an chóras fuaimeanna agus an córas fónaice. Ní chuidíonn sé go mbaintear úsaid as aibítir an Bhéarla, agus go síltear go minic gur féidir scileanna fónaice an Bhéarla a aistriú go Gaeilge. Is a mhalairt an scéal, agus is féidir leis an easpa tuisceana seo deacracht a chruthú. B'fhearr go dteagascfaí litearthacht sa dá theanga ar bhealach a fhreagraíonn do na córais fuaimeanna agus scríofa féin.

Tá tuiscint ag fás ar an ghá le modh teagaisc don Ghaeilge atá bunaithe ar struchtúir fuaimnithe na teanga [59], ach fós, tá easpa acmhainní ann do dhaltaí agus do mhúinteoirí, chun feasacht foinéolaíochta a chothú mar dhúshraith don fheasacht fónaice. Cé go bhfuil eisceachtaí le nótáil [60, 61], is ar an Bhéarla atá formhór na n-ábhar teagaisc don Ghaeilge bunaithe, rud nach n-oireann ar chor ar bith [62]. Dóibh siúd le disléicse, níl uirlisí ar fáil chun measúnú a dhéanamh, nó ábhair tacaíochta ann le tacú le sealbhú na litearthachta (tá uirlisí agus ábhair tacaíochta an Bhéarla iomlán mí-oiriúnach) agus is beag treoir atá ann do mhúinteoirí tacaíochta foghlama agus siceolaíthe [63, 64].

Aithnítear sa Pholasaí don Oideachas Gaeltachta [65] an gá atá leis na hacmhainní seo don Ghaeilge. Caithfear iad a dhearadh as an nua agus caithfear iad a bhunú ar fheasacht ar struchtúir fóineolaíochta agus fónaice na Gaeilge, seachas a bheith ag iarraidh aistriú a dhéanamh ar acmhainní a cumadh don Bhéarla. Tá buntaighde a thacaíonn leis an aidhm seo ar siúl i gColáiste na Trínóide , ina bhfuil fiosrú á dhéanamh ar shealbhú feasachta fóineolaíochta agus fónaice i measc cainteoirí dúchais agus foghlaimeoirí [66-69]. Ceangailte leis seo, tá comhoibriú ABAIR -CabairE ag déanamh forbartha ar chluichí idirghníomhacha a dhíríonn ar an fheasacht fóineolaíochta maraon le bunscileanna litearthachta [70] (féach Caibidlí 12 agus 15). Tá géarghá le háiseanna le freastal ar riachtanais daltaí le disléicse (féach Caibidil 15). Sa todhchaí, ba cheart go bhféadfaí scagthástáil a chur ar fáil ar líne, chomh maith le hábhair oiliúna agus feabhais, a bhaineann leas as na hacmhainní teangeolais agus na háiseanna idirghníomhacha a bheidh ag teacht chun cinn.

Mar achoimre, cuirfidh an teicneolaíocht le sealbhú, teagasc agus measúnú na Gaeilge trí (1) úsáid na n-acmhainní agus uirlisí digiteacha chun soiléiriú a dhéanamh ar na spriocanna tábhachtacha i sealbhú na Gaeilge mar theanga dhúchais nó mar an dara teanga, agus trí (2) fhorbairt ar áiseanna ríomhchuidithe ina bhfuil aird ar na spriocanna, a thacóidh le gach foghlaimeoir – iad siúd le fadhbanna labhartha, teanga agus litearthachta san áireamh.

2.8 Malairt Teanga agus Athrú Teanga

Cosúil le gach córas beo, tá éagsúlacht shuntasach le sonrú sna canúintí ar fad, sna grúpaí sóisialta agus sna haoisghrúpaí cainteoirí, agus i gcomhthéacs na húsáide. Is cinnte go dtiocfaidh athrú ar an teanga le himeacht ama freisin – agus is minic a léiríonn éagsúlacht chanúna na linne seo athruithe éagsúla atá ag tarlú ag rátaí éagsúla in áiteanna geografacha éagsúla. Is minic a bhíonn teagmháil le teanga eile ina foinse d'athrú teanga. Tá impleachtaí ag teagmháil teanga go nádúrtha maidir le sealbhú na Gaeilge mar chéad teanga agus mar dhara teanga, ós rud é gur mionteanga í an Ghaeilge a mhaireann cóngarach go leor do theanga dhomhanda cosúil leis an mBéarla. Is féidir taighde a dhéanamh ar éagsúlacht teanga agus ar athrú teanga trí chorpais chomhaimseartha agus stairiúla Ghaeilge a úsáid (Caibidil 3.6), idir corpais scríofa agus corpais labhartha.

Is gnách go n-athraíonn an teanga scríofa níos lú ná an teanga labhartha, ós rud é go bhfuil caighdeán scríofa bunaithe ann, agus ós rud é go mbíonn claonadh ag an teanga scríofa a bheith coimeádach maidir le gnásanna labhartha atá ag athrú. Mar sin féin, i ndomhan atá ag éirí níos éagsúla ar líne, tá sé tábhachtach staidéar a dhéanamh ar fhoirmeacha scríofa níos athraithí freisin, mar shampla an teanga a úsáidtear sna meáin shóisialta i saol laethúil úsáideoirí na Gaeilge, chomh maith leis an téacs eagarthóireachta atá le fáil sa litríocht nó sa nuacht.

Ba chóir taighde ar éagsúlacht teanga agus ar athrú teanga a chur san áireamh i mbeartais um chaighdeánú teanga, i bpleanáil teanga, i gcruthú gramadaí tagartha agus i gcaighdeáin scríbhneoireachta. Cuireann na hacmhainní teanga seo go léir le teicneolaíochtaí digiteacha ar nós litreoirí, seiceálaithe gramadaí (Caibidil 13), agus feidhmchláir oideachais (Caibidil 14).

Mar a pléadh i gCuid 2.4 thuas, tá éagsúlacht teanga ina ceist lárnach i ndearadh na teicneolaíochta urlabhra. Tá taighde ar éagsúlacht traschanúna riachtanach i gcás teicneolaíochtaí canúna cuí, lena n-áirítear sintéis urlabhra (Caibidil 8), aithint urlabhra (Caibidil 9) agus córais idirphlé

labhartha (Caibidil 10) chomh maith le feidhmchláir atá bunaithe orthu. Tá cineálacha éagsúlachta eile, amhail aois agus inscne, bunriachtanach freisin i ndearadh teicneolaíochtaí agus feidhmchláir atá bunaithe ar chaint.

2.9 Moltaí

1. Moltar go dtabharfaí faoi réimse leathan de thaighde eimpíreach agus ríomhaireachtúil teangeolaíochta ar ghnéithe éagsúla de struchtúr na Gaeilge, ag díriú ar an teanga labhartha agus scríofa. Moltar tosaíocht a thabhairt (i) do ghnéithe atá tábhachtach don teicneolaíocht ina bhfuil bearnaí eolais faoi láthair, agus (ii) do thaighde ar shealbhú na Gaeilge mar an chéad agus an dara teanga.
2. Moltar go mbaintear úsáid as modhanna agus teicnící taighde a chinntíonn go mbeidh aschuir na taighde
   a. mar bhunús do na hacmhainní atá le forbairt i gCuid I
   b. inúsáidte i dtógáil na mbun-teicneolaíochtaí i gCuid II
c. i bhfoirm a cheadaíonn an fheasacht teanga a leabú in áiseanna i gCuid III a thacaíonn le foghlaimeoirí; leo siúd le deacrachtaí urlabhra, teanga nó litearthachta; le pobal na Gaeilge i gcoitinne.
3. Chun 1 agus 2 a chur ar aghaidh, moltar taighde idirdhisciplíneach a chothú, ina mbíonn teangeolaithe, innealtóirí agus ríomheolaithe ag obair as lámha a chéile i soláthar na n-acmhainní teanga don teicneolaíocht, agus ag fiosrú le chéile na bealaí is fearr le feasacht na teanga a leabú in áiseanna teicneolaíochta.
4. Oiliúint ar thaighdeoirí don Phlean Dhigiteach: Tá cúrsaí céime agus iarchéime de dhíth ina gcuirfear oiliúint thrasdhisciplíneach ar thaighdeoirí, sa dóigh is go bhfuil na scileanna cuí acu sa teanga (An Ghaeilge), sa teangeolas agus i réimsí na teicneolaíochta. Ina leithéid de chúrsa, béidh oiliúint ar mhodhanna taighde eimpíreacha digiteacha de dhíth, agus tuiscint ar an teicneolaíocht ar a bhfuil siad ag díriú. Is féidir cúrsaí oiriúnacha a sholáthar trí leathnú ar chúrsaí atá ar fáil cheana féin , chun go bhfreagraíonn siad ar bhealach níos iomláine don Phlean Dhigiteach, nó is féidir cúrsaí úra a bhunú mar is gá.
5. Moltar leanúnachas taighde a chothú trí mhaoiniú ar scoláirí (iarchéimithe agus nuair is féidir, iardhochtúirí) chun earnáil láidir teicneolaíochta don Ghaeilge á thógáil agus a dhaingniú.
6. Béidh gá le forbairt ar chorpais atá feiliúnach don taighde teangeolaíochta agus d'fhorbairt na teicneolaíochta. Dó seo moltar go ndéantar (i) uasdátú ar na corpais atá ann faoi láthair agus iad a chur ar fáil ar bhealach atá éasca le húsáid, agus (ii) forbairt ar chorpais breise (féach fosta Caibidil 3).

Tagairtí

1. Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet (1999). Cambridge, Cambridge University Press, 111-116.

2. Ní Chasaide, A. (1999). Irish, in the Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet, Cambridge, Cambridge University Press, 111-116.

3. Ó Raghallaigh (2014). Fuaimeanna na Gaeilge. Cois Life.

4. Quiggin, E. C. (1906). A Dialect of Donegal. Cambridge University Press

5. Sommerfelt, A. (1922). The Dialect of Torr, Co Donegal. Christania:Jacob Dybwad

6. Ó Cuív, B., (1944). The Irish of West Muskerry, Co. Cork: a phonetic study. Dublin Institute of Advanced Studies

7. de Bhaldraithe, T. (1945). The Irish of Cois Fhairrge, Co. Galway: a phonetic study. Dublin Institute of Advanced Studies.

8. Breatnach, R. B., (1947). The Irish of Ring, Co. Waterford: a phonetic study. Dublin Institute of Advanced Studies.

9. de Búrca, S. (1958). The Irish of Tourmakeady, Co. Mayo: a phonemic study. Dublin Institute of Advanced Studies

10. Mhac an Fhailigh, É., (1968). The Irish of Erris, Co. Mayo: a phonemic study. Dublin Institute of Advanced Studies.

11. Ó Curnáin, B., (2007). The Irish of Iorras Aithneach, County Galway. Dublin Institute of Advanced Studies.

12. Ó Sé, D. (2000). Gaeilge Choirce Dhuibhne. Dublin Institute of Advanced Studies.

13. Bennett, R., G. McGuire, M. Ní Chiosáin & J. Padgett. (2017). An ultrasound study of Connemara Irish palatalization and velarization, in Journal of the International Phonetic Association, published online by Cambridge University Press, pp. 261-304.

14. Padgett, J. and Ní Chiosáin, M.(2018). The perception of a secondary palatalization contrast: A preliminary comparison of Russian and Irish, in Hana-bana (花々): A Festschrift for Junko Ito and Armin Mester, published online.

15. Ní Chiosáin, M. & J. Padgett, 2012. An acoustic and perceptual study of Connemara Irish palatalization. Journal of the International Phonetic Association 42.2, 171-191.

16. Fitzpatrick, L. (2000) Analysis of Irish Lingual Articulation Using

Electropalatography and Electromagnetic Articulography, Ph.D Dissertation, Trinity College Dublin.

17. Ní Chasaide, A. and Fitzpatrick, L. (1995). Assimilation of Irish velarised and palatalised stops, in Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm, Royal Institute of Technology and Stockholm University, Vol. 1, 334-337.

18. Ní Chasaide, A. and Fealy, G. (1991). Articulatory and Acoustic Measurements of Coarticulation in Irish (Gaelic) Stops, in Proceedings of the 12thInternational Congress of Phonetic Sciences, Aix-en-Provence.

19. Ní Chasaide, A. (1977). Acoustic study of the laterals in Donegal Irish and Hiberno-English, M.A Dissertation, University of North Wales, Bangor.

20. Ní Sheighin, M. (2001) An Electropalatographic Study of the Consonants of the Dialect of Irish of Dú Chaocháin, Co Mayo. M.Phil. Dissertation, Trinity College Dublin

21. Russell, É. (2001) An EPG Study of the Irish Dialect of Coirce Dhuibhne. M.Phil. Dissertation, Trinity College Dublin.

22. Cooper, A., (1994). An EPG Study of the Irish Consonants of Cois Fharraige. M.Phil. Dissertation, Trinity College Dublin.

23. Blankenhorn, V., (1981). Intonation in Connemara Irish: a preliminary study of kinetic glides, Studia Celtica Vol. xvi-xvii, 259-79.

24. Bondaruk, A., (2004). The inventory of nuclear tones in Connemara Irish, in Journal of Celtic Linguistics 8, 15–47.

25. Dalton, M. (2008). The phonetics and phonology of the intonation of Irish dialects. PhD Dissertation, Trinity College, Dublin.

26. Dalton, M. and Ní Chasaide, A.(2007). Nuclear accents in four Irish (Gaelic) dialects. Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrücken, 965-968.

27. Dalton, M. and Ní Chasaide, A. (2005). Tonal alignment in Irish Dialects. Language and Speech, 48 (4), pp. 441-464.

28. Dorn, A., O'Reilly, M. and Ní Chasaide, A. (2011). Prosodic signalling of sentence mode in two varieties of Irish (Gaelic). Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, 611-614.

29. Dorn, A. and Ní Chasaide, A. (2015). Sentence mode differentiation in four Donegal Irish varieties. Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow.

30. O'Reilly, M. and Ní Chasaide, A. (2016). Modelling the timing and scaling of nuclear pitch accents of Connaught and Ulster Irish with the Fujisaki model of intonation. Proceedings of the 8th Int.Conference on Speech Prosody, Boston, pp. 355–359.

31. Dalton, M. & Ní Chasaide, A. (2007). Melodic alignment and micro-dialect variation in Connaught Irish. In C. Gussenhoven & T. Riad (Eds.), Tones and Tunes: Studies in Word and Sentence Prosody, Vol. 2, Mouton de Gruyter, Berlin, 293-315.

32. Ní Chasaide, A., Ní Chiaráin, N., Berthelsen, H., Wendler, C., Murphy, A., Barnes, E. and Gobl, C. (2019). Leveraging phonetic and speech research for Irish language revitalisation and maintenance. Proceedings of the 19th Int.Congress of Phonetic Sciences, Melbourne, Australia, pp. 994-998.

33. Na Bráithre Críostaí, Graiméar Gaeilge na mBráithre Críostaí. 1960; 1999, Baile Átha Cliath: An Gúm.

34. Christian Brothers, New Irish Grammar. 1988, Dublin: C J Fallon.

35. Uí Dhonnchadha, E., C. Nic Pháidín, and J. van Genabith, Design, Implementation and Evaluation of an Inflectional Morphology Finite-State Transducer for Irish. MT Journal - Special Issue on Finite State Language Resources and Language Processing, 2005.

36. Rannóg an Aistriúcháin, Gramadach na Gaeilge agus Litriú na Gaeilge: An Caighdeán Oifigiúil. 1958, Baile Átha Cliath: Oifig an tSoláthair.

37. Carnie, A. and E. Guilfoyle, eds. The Syntax of Verb Initial Languages. 2000, Oxford University Press: Oxford.

38. Duffield, N., Particles and Projections in Irish Syntax. 1995, Dordrecht: Kluwer.

39. McCloskey, J., A VP in a VSO language?, in Order, Concord and Constituency, G. Gazdar, E. Klein, and G. Pullum, Editors. 1983, Foris: Dordrecht. p. 9-55.

40. Ó Donnchadha, G., Syntactic Structure Building and the Verbal Noun in Modern Irish: A Minimalist Approach. 2010, University College Dublin: Dublin.

41. Stenson, N., Studies in Irish Syntax. Ars Linguistica, ed. W. Werner. 1981, Tübingen: Gunter Narr Verlag.

42. Nolan, B., The Structure of Modern Irish: A functional approach. 2012: Equinox.

43. Bresnan, J., et al., Lexical Functional Syntax. 2016: Wiley Blackwell.

44. Pollard, C. and I.A. Sag, Head-Driven Phrase Structure Grammar. 1994, Chicago: University of Chicago Press.

45. Asudeh, A. The Syntax of Preverbal Particles and Adjunction in Irish. in LFG02 Conference. 2002. Athens: CSLI Publications.

46. Sulger, S., Irish Clefting - The LFG perspective. 2009, Konstanz.

47. Mel'čuk, I.A., Dependency Syntax: Theory and Practice. 1988, Albany: State university of New York Press.

48. Harrington, S., Singleton, D. & Henry, A. (2000). At the sharp end of language revival: English-speaking parents raising Irish-speaking children. In CLCS Occasional Paper 57, 12 pp.. Trinity College Dublin

49. Hickey, T. (2012). ILARSP: A Grammatical Profile of Irish. In Ball, M., Crystal, D. and Fletcher, P. (eds.). The Languages of LARSP. Multilingual Matters

50. Goodluck, H, Guilfoyle, E. & Harrington, S. (2001). Acquiring Subject and Object Relatives: Evidence from Irish. In Kallen, J. (Ed.) Journal of Celtic Language Learning, Vol 6 First Language Learning.

51. Brennan, S. (2004). Na chéad chéimeanna : luathfhorbairt Gaeilge mar phríomhtheanga : díriú ar an bhfóineolaíocht [First steps : early development of Irish as a primary language : focus on phonology]. Western Health Board

52. Ní Ghloinn, A., E. Uí Dhonnchadha, and A. O'Keeffe. The design and annotation of the TEG learner corpus of Irish. in Inter-Varietal Applied Corpus Studies International Biennial Conference. 2018. Malta.

53. Council of Europe, Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). 2001: https://rm.coe.int/1680459f97.

54. Council of Europe Common European Framework of Reference for Languages: Learning, teaching, assessment. Structured overview of all CEFR scales. 2011.

55. Council of Europe, Common European Framework of Reference For Languages: Learning, Teaching, Assessment - companion Volume with New Descriptors. 2018: https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989.

56. Ó Meachair, M.J., The Creation and Complexity Analysis of a Corpus of Educational Materials in Irish (EduGA). 2019, Trinity College, Dublin. p. 208.

57. Seymour, p., Aro, M., & Erskine, J. (2003). Foundation literacy acquisition in European orthographies, in British Journal of Psychology, 94, 143–17

58. [58] Stenson, N., and Hickey, T. (2014). "In defense of decoding." Journal of Celtic Language Learning 18: 11-40.

59. Stenson, N., & Hickey, T. (2018). Understanding Irish Spelling. An Chomhairle um Oideachas Gaeltachta agus Gaelscolaíochta. Retrieved from: https://www.cogg.ie/wp-content/uploads/Understanding-Irish-Spelling-A-Handbook-for-Teachers-and-Learners-by-Dr.-Nancy-Stenson-and-Dr.-Tina-Hickey-1.pdf

60. Breacadh: Mar a Déarfá, Casla, Co. na Gallaimhe.

61. Education and Library Board, Fónaic na Gaeilge. (2011), Béal Féirste

62. Nic Pháidín, F. (2017). Assessing Irish literacy resources: guidelines for teacher and publishers. In Exploring the Literacy Landscape: Proceedings of the 40th Conference of the Literacy Association of Ireland.

63. Barnes, E. (2017). Dyslexia assessment and reading intervention for pupils in Irish medium education: Insights into current practice and consideration for improvement. (Unpublished Master's thesis).

64. Barrett, M. (2016). Doras feasa fiafraí: Exploring special educational needs provision and practices across Gaelscoileanna and Gaeltacht primary schools in the Republic of Ireland. (Unpublished Master's thesis). Retrieved from: http://www.cogg.ie/wp-content/uploads/doras-feas-fiafrai.pdf

65. An Roinn Oideachais agus Scileanna, Polasaí don Oideachas Gaeltachta 2017-2022. 2016: https://www.education.ie/ga/Foilseach%C3%A1in/Tuarasc%C3%A1lacha-Beartais/Polasai-don-Oideachas-Gaeltachta-2017-2022.pdf.

66. Barnes, E., Ní Chasaide, A., Ní Chiaráin, N. (2020). Bilingual phonological awareness: when interdependence becomes interference, in Proceedings of the 15th Congress of the International Association for the Study of Child Language, Philadelphia.

67. Barnes, E., Ní Chasaide, A. and Ní Chiaráin, N. (2018). The design and pre-testing of literacy and cognitive tasks in Irish and English. Proceedings of the Literacy Association of Ireland 42nd International Conference, Dublin, Ireland, pp. 1-11.

68. Barnes, E., Ní Chiaráin, N. and Ní Chasaide, A. (2017). Departures from the "norm": how the phonology, morphology and orthography of the Irish language impact on literacy instruction and acquisition, in Exploring the Literacy Landscape: Celebrating 40 Years of Research and Practice. Literacy Association of Ireland 40th Annual Conference, Marino Institute of Education, Dublin, B. Culligan and G. Mehigan (eds.), pp. 22-32.

69. Ní Chasaide, A., Ní Chiaráin, N., Berthelsen, H., Wendler, C., Murphy, A., Barnes, E. and Gobl, C. (2019). Leveraging phonetic and speech research for Irish language revitalisation and maintenance. Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, pp. 994-998.

Caibidil 3
Acmhainní Teanga: Corpais

3.1 Céard is Corpas ann?

Is éard is corpas ann ná bailiúchán de shamplaí teanga i bhformáid dhigiteach ina dtaispeántar cén chaoi a n-úsáidtear teanga go praiticiúil. Féadfaidh na samplaí teanga seo a bheith i bhfoirm bailiúcháin téacsanna, taifeadtaí den teanga labhartha, agus tras-scríbhinní (tras-scríbhinní foghraíochta nó ortagrafacha) d'ábhar fuaime nó físe.

Is éard atá i gcorpais áirithe ná bailiúcháin mhóra ghinearálta. Is gnách go bhfaightear iad seo ó shamplaí teanga a tharlaíonn go nádúrtha nuair is féidir, ach i mórán cásanna is gá na corpais a chruthú chun críocha sonracha. I measc na samplaí tá corpais urlabhra atá deartha do shintéis urlabhra, corpais foghlaimeoirí atá deartha chun na cineálacha earráidí agus gnéithe eile de theanga foghlaimeoirí a fhiosrú, nó corpais meán sóisialta atá deartha chun léargas a thabhairt ar úsáid teanga ar líne.

3.2 Cén fáth agus cé dó a bhfuil corpais tábhachtach?

Tá corpais riachtanach do theicneolaíochtaí urlabhra agus teanga mar gheall go gcuireann siad eolas ar fáil ar ghnáthúsáid teanga atá ag teastáil chun brí a bhaint as teanga scríofa agus labhartha agus í a ghiniúint. Teastaíonn corpais chun na hacmhainní, na bunachair eolais agus na múnlaí at leagtha amach i gCuid I a thógáil. Múnlaíonn an fhianaise eimpíreach a fhaightear ó chorpas an fhaisnéis theangeolaíoch a fhaighimid i bhfoclóirí, i leabhair ghramadaí agus i dteasárais. Leis na corpais, léirítear an teanga i gcomhthéacs (is é sin i gcomhshuíomhanna), éascaítear ríomh a dhéanamh ar mhinicíocht chomparáideach na bhfocal agus na struchtúr araon, agus cuirtear eolas luachmhar ar fáil ar gach gné de struchtúr teanga – foghraíocht-fóineolaíocht, comhréir, séimeantaic agus pragmataic.

Taobh amuigh d'acmhainní a sholáthar, úsáidtear corpais go díreach chun na bunteicneolaíochtaí atá leagtha amach i gCuid II a thógáil agus a thástáil, mar shampla sintéis urlabhra, aithint cainte agus meaisínaistriúchán. I gcásanna áirithe teastaíonn corpas mar chuid dhílis den chóras ag am rite (mar shampla sintéis roghnúcháin aonad (Caibidil 8) agus meaisínaistriúchán (Caibidil 11)). I saol seo an olltiomsaithe sonraí agus na dteicneolaíochtaí faoi thionchar staitisticí nó néarbhunaithe ina n-úsáidtear líonta néaracha doimhne, tá sé ríthábhachtach go mbeadh fáil ar líon mór corpas. I gcás go leor de na feidhmchláir a shamhlaítear i gCuid III, tá ról tábhachtach ag corpais freisin. Mar shampla, cuireann faisnéis ó chorpais foghlaimeoirí, in éineacht leis an eolas teangeolaíoch (a thagann ó chineálacha eile corpas), ionchur riachtanach ar fáil sna feidhmchláir oideachais CALL den chineál is déanaí.

Le blianta beaga anuas, tá na forbairtí agus an dul chun cinn is suntasaí a rinneadh i réimse PTN (Caibidil 5) bunaithe ar úsáid samhlacha leabaithe focal agus claochladáin. Is cineál cur chuige néarlíonra iad seo a fhoghlaimíonn ar bhealach níos cruinne faoi nádúr na teanga trí chomhthéacs na bhfocal a mheas. Tá tionchar le feiceáil cheana féin ag an gcéad mhúnla claochladáin don Ghaeilge – gaBERT [4] – atá traenáilte ar an méid teoranta corpas atá ann cheana féin – ar chaighdeán na n-uirlisí reatha PTN. Dá bhrí sin, tá an dlúthcheangal idir infhaighteacht na gcorpas agus caighdeán na n-uirlisí teicneolaíochta teanga ag éirí níos soiléire, agus ba chóir go mbeadh an bhéim is mór air sin ó thaobh na Gaeilge de feasta.

Mar sin rachaidh corpais chun leasa an phobail ar go leor bealaí. Bainfidh an príomhbhuntáiste go hindíreach as na teicneolaíochtaí urlabhra agus teanga iad féin, fiú nuair nach mbíonn a bhfuil sa chorpas le feiceáil ag an úsáideoir. I gcásanna áirithe, déanfar corpais a óstáil le comhéadain inchuardaithe ar shuíomhanna idirlín poiblí ar féidir le húsáideoirí rochtain dhíreach a fháil orthu chun a gcríoch féin.

3.3 Cineálacha Corpas

D'fhéadfadh samplaí de theanga scríofa nó labhartha nó an dá rud a bheith i gcorpas. Stóráiltear na samplaí teanga sin ar ríomhairí mar chomhaid téacs agus mar chomhaid fuaime/físe.

Is éard a bhíonn i gcorpais scríofa go hiondúil ná samplaí teanga a tógadh ó fhoinsí foilsithe arna gcur in eagar, amhail leabhair, ailt nuachtáin, tuarascálacha, lámhleabhair agus téacsleabhair. Cuirtear raon leathan topaicí agus seánraí san áireamh de ghnáth ar mhaithe le hionadaíocht teanga a bheith ann. Is gnách go mbíonn na corpais scríofa seo ceart ó thaobh na gramadaí

agus úsáidtear litriú caighdeánach iontu, agus mar sin cuireann siad múnla idéalach den teanga ar fáil. Le déanaí, mar gheall ar an méadú ar úsáid an idirlín agus na meán sóisialta, tá gá le feidhmchláir teanga atá in ann ábhar a ghintear ag úsáideoirí (ÁGÚ) a láimhseáil, m.sh. (ó ardáin meán sóisialta ar nós Twitter nó Facebook,) rud nach mbíonn chomh cruinn céanna ó thaobh na gramadaí agus féadfaidh litriú neamhchaighdeánach agus giorrúcháin a bheith ann a chruthaíonn dúshláin d'uirlisí próiseála. Le ÁGÚ tugtar léargas ar an teanga in úsáid laethúil na meán sóisialta agus úsáidtear corpais dá leithéid chun críocha sochtheangeolaíochta agus PTN araon. Bíonn corpais scríofa le feiceáil i bhformáid amhthéacs, agus anuas air sin d'fhéadfadh anótálacha úsáideacha teangeolaíochta a bheith iontu freisin ar féidir iad a dhéanamh de láimh, go leath-uathoibríoch nó go huathoibríoch (féach mír 3.4). Tá luach ar leith ag baint leis na corpais seo i bhforbairt uirlisí lena bpróiseáiltear teanga daoine go huathoibríoch (féach Caibidil 5).

Féadfaidh corpais labhartha le tras-scríbhinní ortagrafacha nó foghraíochta faisnéis bhreise riachtanach a sholáthar chun cur leis an méid is féidir a fháil ó chorpas téacs amháin. Is éard atá iontu seo ná bailiúcháin leathana ghinearálta de shamplaí teanga fuaime/físe, bídís spontáineach nó scriptithe, ag brath ar chuspóir an chorpais. Úsáidtear samplaí teanga spontáineacha chun feiniméin nach bhfuil chomh coitianta céanna sa teanga scríofa a dhoiciméadú, m.sh. gnéithe dioscúrsa, míreanna léacsacha áirithe, moilleadóireacht, athrá, athcheapadh abairtí, abairtí neamhiomlána, srl. Déantar na tras-scríbhinní ortagrafacha/foghraíochta a shioncronú leis an gcomhad fuaime ar mhaithe le cuardach feabhsaithe agus tuilleadh próiseála cainte/teanga. Ar mhaithe le hionadaíocht, is gnách go mbíonn raon leathan cainteoirí i raon leathan suíomhanna nádúraíocha i gcorpais labhartha spontáineacha. Tá corpais labhartha riachtanach do dhoiciméadú teanga agus d'fheidhmchláir foghlama teanga.

Tá an-éagsúlacht i measc corpais chainte ina ndírítear ar theicneolaíocht chainte, idir chorpais labhartha ginearálta den chineál sin (ó chaint spontáineach nó ó chaint léite) agus chaint leath-spontáineach (cásanna ina bhfuil an chaint spontáineach ach an t-ábhar srianta) agus léamh ábhar atá deartha go han-cháiréiseach. Toisc nach bhfuil aon fhoirm 'chaighdeánach' labhartha Gaeilge amháin ann, tá sé tábhachtach go gcuirfí corpais ar fáil do na canúintí éagsúla. Cé gur taifeadtaí fuaime den chuid is mó atá i gcorpais chainte, féadfar comharthaí urlabhraíochta nó aeraidinimiciúla a chur san áireamh de réir mar is cuí. Léiríonn éagsúlacht na gcorpas éagsúlacht na dteicneolaíochtaí a bhfuil siad ceaptha dóibh. Mar sin, pléitear na cineálacha éagsúla corpas i gcomhthéacs na dteicneolaíochtaí seo i gCaibidlí 8, 9, agus 10 agus déantar achoimre ar roinnt de na príomhghnéithe anseo. Maidir le sintéis chainte (Caibidil 8) tá cáilíocht ghuth an chainteora agus cáilíocht an taifeadta fuaime thar a bheith tábhachtach. Déantar taifeadtaí atá 'glan' i stiúideo ardsonraíochta, go hiondúil (cé nach gá) d'ábhair ullmhaithe arna léamh ag duine aonair (a roghnaítear go cáiréiseach). Maidir le haithint cainte, chomh maith le corpais ghinearálta an-mhór, lena n-áirítear go leor guthanna, canúintí, agus coinníollacha taifeadta, teastaíonn corpais a mheaitseálann na húsáideoirí agus na feidhmchláir bheartaithe dá bhfuil an córas ceaptha, m.sh. leanaí i suíomhanna oideachais (féach Caibidil 9). Ar an gcaoi chéanna, teastaíonn corpais chainte chomhráiteacha (idirghníomhacha) ó chórais éifeachtacha dialóige (Caibidil 10) ionas gur féidir samhaltú a dhéanamh ar dhialóga fíorshaoil. Maidir le hanailís teangeolaíochta a dhéanamh ar phrosóid chainte Ghaeilge (tuin chainte, malairt shéiseach, aiceannú) agus ar urlabhraíocht (i gCaibidil 2.4) agus a samhaltú do theicneolaíochtaí cainte (Caibidil 7), teastaíonn corpais atá deartha go cáiréiseach agus ina léirítear na feiniméin bheachta spéise, agus ag an am céanna deireadh a chur le tionchar na n-athróg eile.

3.4 Corpais a Ullmhú agus a Anótáil

Is é an chéad tasc a bhaineann le corpas a chruthú tar éis dearadh an chorpais a bheith tugtha chun críche ná na samplaí teanga a bhailiú. Ansin is minic a dhéantar anótáil nó lipéadú ar chorpais le faisnéis teanga. I gcás comhaid fuaime labhartha, féadfaidh sé go mbeadh gá leis na taifeadtaí a thras-scríobh, go hortagrafach ar a laghad (ach féach tuilleadh thíos).

3.4.1 Corpais Scríofa

I gcás téacsanna scríofa atá i bhformáid dhigiteach cheana féin, is gnách go mbíonn gá le 'glanadh' forleathan chun ábhar nach dteastaíonn a bhaint astu. Bíonn corpais ghnáth-théacs an-úsáideach ach de ghnáth ní bhíonn ann ach an chéad chéim. Don chuid is mó de na cásanna, déantar an fhaisnéis is díol spéise d'fheidhmchlár a anótáil go soiléir (lipéadaithe, clibeáilte) sna sonraí. Cuirtear anótálacha struchtúracha agus teanga araon i bhfeidhm go coitianta. Léirítear i bhFíor 3.1 sampla de ghnáth-théacs ina dtaispeántar go soiléir struchtúr anótálacha struchtúir paragraf <p> agus abairt <s> do na sonraí: Bhí Máire fuar 'Mary was cold'. Chuir sí geansaí olla uirthi 'She put a woollen jumper on her'.

```
<p>
<s> Bhí Máire fuar. </s><s> Chuir sí geansaí olla uirthi. </s>
</p>
```

Fíor 3.1: gnáth-théacs maille le marcáil struchtúrach paragraif agus abairtí

Léirítear i bhFíor 3.2 abairt ghearr ina gcuimsítear anótálacha struchtúracha agus teangeolaíocha. Sa sampla seo, tá gach focal anótáilte go soiléir lena leama (ceannfhocal an fhoclóra) agus le clib a thaispeánann an roinn cainte (briathar, ainmfhocal srl.) agus faisnéis

faoi fheidhm ghramadaí (ainmní, cuspóir, srl.) agus caidreamh spleáchais-cinn (m.sh. #4->3; tá comhartha 4 ag brath ar cheann 3). Tá lipéadú den chineál sin bunúsach don chuid is mó de na huirlisí agus de na feidhmchláir a gcuirtear síos orthu i gCaibidil 5.

```
<p><s>
<w lemma="cuir" tag="Verb+PastInd@FMV#1->1">Chuir</w>
<w lemma="Máire" tag="Prop+Noun+Fem+Com+Sg@SUBJ#2->1">Máire</w>
<w lemma="geansaí" tag="Noun+Masc+Com+Sg@OBJ#3->1">geansaí</w>
<w lemma="olann" tag="Noun+Fem+Gen+Sg@N<#4->3<">olla</w>
<w lemma="ar" tag="Pron+Prep+3P+Sg+Fem@PP_OBL#5->1">uirthi</w>
<w lemma="." tag="Punct+Final#6->6">.</w>
</s></p>
```

Fíor 3.2: Abairt atá anótáilte le leama, roinn cainte, feidhm ghramadaí agus gaol spleáchais–cinn do "Chuir Máire geansaí olla uirthi"

Is éard atá i mbunachar crainn ná corpas téacs atá anótáilte le faisnéis maidir le struchtúr comhréire an téacs (Fíor 3.3). Tagraítear do chorpas mar seo mar bhunachair chrainn toisc go léirítear abairtí mar chrainn pharsála, ina nasctar focail agus nathanna cainte trí naisc atá cosúil le craobhacha a chuireann síos ar ghaolmhaireachtaí agus ar struchtúir ghramadaí (féach Aguisín A). Tá gá le bunachair chrainn le haghaidh taighde comhréire teangeolaíoch agus chun parsálaithe comhréire atá bunaithe ar staitisticí a oiliúint (féach Caibidil 5).

```
# sent_id = 904
# text = Creidtear gur go mailíseach a tosaíodh an tine.
1 Creidtear   creid          VERB   VTI      Mood=Ind|Tense=Pres|Voice=Auto          0   root
2 gur         is             AUX    Cop      Tense=Pres|VerbForm=Cop                 4   cop
3 go          go             PART   Ad       PartType=Ad                             4   mark:prt
4 mailíseach  mailíseach ADJ Adj    Degree=Pos                                       1   ccomp
5 a           a              PART   Vb       PartType=Vb|PronType=Rel                6   mark:prt
6 tosaíodh    tosaigh        VERB   VTI      Mood=Ind|Tense=Past|Voice=Auto          4   csubj:cleft
7 an          an             DET    Art      Definite=Def|Number=Sing|PronType=Art   8   det
8 tine        tine           NOUN   Noun     Case=NomAcc|Gender=Fem|Number=Sing      6   obj
9 .           .              PUNCT  .        _                                       1   punct
```

Fíor 3.3: Abairt atá anótáilte i mBunachar Crainn Spleáchas Uilíoch Gaeilge, ina bhfuil: uimhir aitheantais na habairte, téacs amh, leama, roinn cainte, gnéithe moirfeolaíocha, mar aon le lipéid struchtúir gramadaí agus ceangaltáin

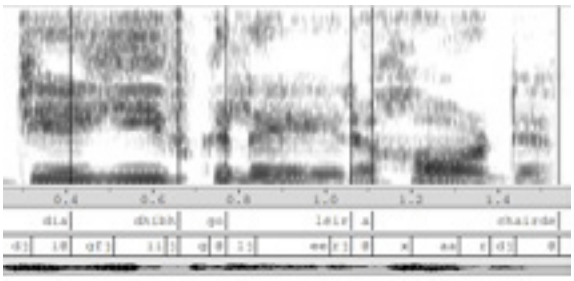
Is éard atá i gcorpais atá anótáilte go séimeantach ná corpais ina bhfuil anótáil shéimeantach maidir le míreanna léacsacha aonair agus struchtúir ghramadaí a léirítear i bhFíor 3.4. Bunachar tairiscintí a thugtar ar chorpas ina bhfuil argóintí briathra agus faisnéis eile anótáilte, mar shampla PropBank (1) (féach Caibidil 2, séimeantaic). Déantar anótáil ar gach briathar i gcorpas roghnaithe, rud a thugann clúdach leathan ar úsáid agus ar éagsúlacht briathra. Is féidir aonáin ainmnithe a lipéadú freisin, rud a chuidíonn le cineálacha nathanna cainte ainmfhoclacha a idirdhealú agus a róil shéimeantacha a chinneadh. Tá an fhaisnéis seo riachtanach le haghaidh tascanna iartheachtacha amhail aisghabháil faisnéise ardchaighdeáin, achoimriú téacs agus meaisínaistriúchán.

```
[Mr.  Bush] met [him] [privately], [in the White House], [on Thursday].
 ARG0      REL  ARG1  ARGM-MNR     ARGM-LOC            ARGM-TMP
```

Fíor 3.4: Abairt anótáilte go séimeantach bunaithe ar shampla i dTreoirlínte Anótála PropBank Béarla (2) ag baint úsáid as lipéadú séimeantach Propbank ARG0: gníomhaí, ARG1: iomfhulangach, ARGM-MNR: cáilitheoir modha, ARGM-LOC: cáilitheoir áitreabhach agus ARGM-TMP: cáilitheoir ama.

3.4.2 Speech Corpora

I gcás corpais chainte atá beartaithe le haghaidh teicneolaíocht chainte, sainaithnítear na fuaimeanna aonair i gcomhad fuaime agus déantar an comhartha a dheighleánú chun a aithint cén áit a dtosaíonn agus a gcríochnaíonn siad. Is éard atá i gceist le seiceáil anótála ná cigireacht amhairc ar an speictreagram cainte (an chuid uachtair d'Fhíor 3.5) agus éisteacht cháiréiseach, lena chinntiú go dtagann an tras-scríobh leis an méid a labhraíodh, go bhfuil an deighleánú i gceart, go n-aithnítear moilleadóireacht agus earráidí cainte, srl.



Fíor 3.5: Sampla ó chorpas cainte anótáilte ABAIR, ina dtaispeántar an speictreagram cainte maille le deighleánú ortagrafach agus foghraíochta agus tras-scríobh (Tá léaráid nua ag teastáil anseo le + waveform + IPA)

D'fhéadfadh leibhéil bhreise anótála teacht ina dhiaidh sin, ag brath ar an gcuspóir a bhfuil an corpas beartaithe dó. Is minic a éilíonn na nótaí seo réamhphróiseáil ar an gcomhartha. Mar shampla, má tá gá le heolas prosóide maidir le séis chainte (mar shampla le cur i bhfeidhm i gcórais sintéise/dialóige labhartha), is féidir le céimeanna próiseála tosaigh ar nós eastóscadh toinne, anótáil a cheadú ina dhiaidh sin ar phointí tábhachtacha toinne, mar shampla pointí arda agus ísle sa chomhrian séiseach. Tríd an gcomhartha cainte a phróiseáil agus trí leibhéil bhreise nótaí mínghráinneacha a chur leis, déantar bunachar saibhir den chorpas – rud atá úsáideach don anailís foghraíochta-theangeolaíoch (Caibidil 2 agus riachtanach do shamhaltú phróisis phrosóide agus urlabhraíochta na Gaeilge, mar a thuairiscítear i gCaibidil 7).

3.4.3 Deighleánú láimhe agus uathoibríoch

Féadfaidh na milliúin focal, nó go leor uaireanta de chaint taifeadta, a bheith i gcorpais. Go hiondúil, dá mhéad an corpas is ea is úsáidí é. Tá infheistíocht shuntasach ama agus saineolais i gceist le hanótáil corpais. Tagraítear do chorpais anótáilte atá seiceáilte agus fíoraithe de láimh (ag beirt saineolaithe nó níos mó) mar chorpais ar chaighdeán óir agus úsáidtear iad go díreach in oiliúint agus meastóireacht a dhéanamh ar uirlisí próiseála cainte/teanga agus i dtógáil teicneolaíochtaí. Ag na céimeanna tosaigh d'ullmhú an chorpais, de ghnáth is de láimh a dhéantar anótáil teanga, próiseas a éilíonn eolas ar an teanga (nó ar an gcanúint) agus ar na gnéithe teangeolaíocha atá á gcur san áireamh. Mar sin féin, is tasc neamhphraiticiúil é corpais mhóra a anótáil de láimh. Má tá corpas sách mór sonraí lámh-anótáilte ar fáil, is féidir na sonraí sin a úsáid, trí mheaisínfhoghlaim, chun an próiseas anótála le haghaidh corpais bhreise a uathoibriú (nó leath-uathoibriú) (tugtar "bútáil" ar an modh seo uaireanta). Is féidir corpais ar chaighdeán óir a fhorbairt níos tapúla ansin trí lipéadú uathoibrithe a cheartú de láimh, seachas sonraí a anótáil de láimh ón tús. Maidir le corpais chainte, is féidir algartaim deighleánúcháin uathoibríocha a úsáid chun ailíniú tosaigh de na fuaimeanna aonair (cur síos foghraíochta) a chur ar fáil don tonnchruth cainte.

3.5 Corpais Speisialaithe

Léirítear sa chuid seo roinnt samplaí de chorpais speisialaithe, a bhaineann go háirithe leis an bPlean Digiteach.

Bíonn idir bhuntéacs agus a n-aistriúcháin araon i gcorpais chomhthreomhara, atá ailínithe ag leibhéal na habairte. Foghlaimíonn meaisínaistriúchán staitistiúil (MAS) agus meaisínaistriúchán néarach (MAN) araon patrúin ó bhailiúcháin mhóra de théacs comhthreomhar chun aistriúcháin a thuar. Mar shampla, i gcorpas comhthreomhar Béarla-Gaeilge bíonn corpas mór de théacs Béarla ailínithe lena choibhéis aistrithe bheacht go Gaeilge. Féadfaidh sé a bheith dúshlánach sonraí maithe comhthreomhara a aimsiú, go háirithe i gcás teangacha atá ar bheagán acmhainní ar nós na Gaeilge. Go minic is é an príomhthasc a bhaineann le corpas comhthreomhar a chruthú ná sonraí a aistríodh roimhe seo ó aistritheoirí nó ó ghnólachtaí aistriúcháin a aithint agus a bhailiú nó dul i dteagmháil le ranna rialtais nó eagraíochtaí leathstáit cuí. Ina theannta sin, úsáidtear uirlisí bogearraí ar a dtugtar 'ransaitheoirí gréasáin' uaireanta chun téacsanna comhthreomhara a aimsiú ar an idirlíon. Tá sé ríthábhachtach ailíniú ceart ar fud na n-abairtí ionas nach bhfoghlaimeoidh an córas MAS aistriúcháin atá mícheart.

Is féidir le corpais inchomparáide a bheith úsáideach freisin le haghaidh an mheaisínaistriúcháin. I gcodarsnacht leis an gcorpas comhthreomhar, is tacair sonraí iad seo le téacs i mBéarla agus a choibhéis garbh i nGaeilge. D'fhéadfadh sé seo éirí as ábhar a ghiniúint ar thopaic sa dá theanga go comhuaineach seachas a bheith ag aistriú ó cheann go chéile. Sampla de seo is ea Vicipéid (féach Aguisín A) agus suíomhanna nuachta ar líne, áit a gcruthaítear ailt ábhair go neamhspleách ar a chéile go minic, ach go gclúdaíonn siad an t-ábhar céanna. Is féidir le corpais inchomparáide léargas saibhir a thabhairt ar na héagsúlachtaí sna nathanna cainte idir teangacha,

rud atá luachmhar don anailís teangeolaíochta agus don mheaisínfhoghlaim araon.

Cuimsíonn corpais foghlaimeoirí samplaí de theanga an fhoghlaimeora a bhailítear ag céimeanna éagsúla den phróiseas foghlama. Is féidir teanga scríofa agus labhartha bheith san áireamh sna samplaí teanga seo. Is acmhainn riachtanach iad corpais foghlaimeoirí chun an próiseas foghlama teanga a thuiscint. Cuireann siad ar ár gcumas freisin gnéithe sainiúla teanga lena mbaineann deacrachtaí ar leith don fhoghlaimeoir a aithint, mar shampla gnéithe atá casta ó nádúr, gnéithe atá neamhchoitianta in úsáid ghinearálta, nó atá an-éagsúil le céadteanga an fhoghlaimeora. Tá an fhianaise eimpíreach, a bhaineann leis an ord agus an ráta ag a ndéantar máistreacht ar ghnéithe ar leith agus na leibhéil deacrachta a chuireann siad romhainn, luachmhar chun curaclaim agus ábhair foghlama teanga a mhionathrú. Soláthraíonn sé seo buntaca bunriachtanach freisin d'fhorbairt teicneolaíochtaí foghlama ríomhchuidithe teangacha a dhíríonn ar shainréimsí deacrachta, agus atá oiriúnach do na céimeanna éagsúla den sealbhú.

Is éard atá i gcorpais a bhaineann go sonrach leis an bhfearann ná bailiúcháin téacs a bhailítear de réir sainréimsí nó seánraí, amhail téacs dlíthiúil, ábhair eolaíochta, ábhair oideachais, seirbhísí nuachta, agus mar sin de. Maidir le huirlisí próiseála teanga atá bunaithe ar staitisticí agus a dhéileálann le cineálacha éagsúla téacs, braitheann siad ar infhaighteacht a leithéid de théacsanna a bhaineann go sonrach leis an bhfearann chun iad a fhorbairt agus a thástáil. Teastaíonn iarracht spriocdhírithe bailithe chun sonraí sonracha den chineál sin a bhailiú. Mar sin féin, is féidir ábhar ó sheirbhísí nuachta agus ábhar ar líne (i gcás ina gceadaítear é) a fháil trí ransú gréasáin.

Corpais chainte (sintéis): teastaíonn corpais chainte fhairsinge le haghaidh gach forbairt teicneolaíochta cainte (luaitear corpais don chaint shintéiseach anseo). Braitheann méid agus comhdhéanamh an chorpais ar an modheolaíocht sintéise a ghlacfar (féach Caibidil 7), ach is cuma cén méid, tá gá le clúdach foghraíochta lena chinntiú go gcuirtear fuaimeanna uile na canúna san áireamh i ngach iomalartú agus comhcheangal ceadaithe. Bíonn clúdach maith i gceist le corpas an-mhór, ach tá sé costasach ó thaobh an mhéid oibre a theastaíonn ó dhaoine le haghaidh anótála, ceartúcháin, srl.

3.6 A bhfuil déanta i leith na Gaeilge go dtí seo

Tá cur síos ar na corpais a forbraíodh don Ghaeilge go dtí seo, nó atá á fhorbairt faoi láthair, i dTáblaí 3.1 agus 3.2. De réir caighdeáin idirnáisiúnta, tá siad seo uile measartha beag agus i ngach cás teastaíonn ábhair bhreise a bhailiú nó a thaifeadadh. Ina theannta sin, ní mór tús a chur le modhanna nua agus feabhsaithe anótála (leath-)uathoibríocha. Tá tuilleadh sonraí ar chuid de na corpais seo le fáil in Aguisín A mar aon le hachoimre ar mholtaí don fhorbairt amach anseo. Tá na hacmhainní sin ar fad agus tuilleadh liostaithe sa Ghreille Eorpach Teangacha, a gcuirfear acmhainní nua léi go rialta.

Tábla 3.1: Corpais scríofa

Is den ríthábhacht é go mbeadh fáil ar chorpais scríofa chun gur féidir teicneolaíochtaí úrscothacha an lae inniu a fhorbairt. Ach de dheasca ceadúnú srianta ar chorpais atá ann cheana, tá bac á chur ar a lán forbairtí dá leithéid. Is fiú a thabhairt faoi deara, mar sin, i dteannta na gcorpas atá liostaithe sa tábla thíos, go bhfuil Corpas Náisiúnta na hÉireann á fhorbairt ó Eanáir 2022 ag taighdeoirí in Fiontar & Scoil na Gaeilge in DCU agus TCD. Is céim shuntasach í an obair seo i dtreo na corpais atá ann cheana féin a chomhlánú le hábhar teanga ón tréimhse le 20 bliain anuas agus ó thaobh aghaidh a thabhairt ar shaincheisteanna inrochtaineachta NCI.

Corpas	Cur síos	Sonraí Anótála
1. Corpas Gaeilge NCI *http://corpas.focloir.ie/*	30 milliún focal de théacs scríofa deartha le bheith cothrom agus ionadaíoch.	Téacschomharthú uathoibrithe bunaithe ar rialacha agus clibeáil de réir na rann cainte.
2. Corpas de Chaighdeán Ór clibeáilte de réir na roinne cainte	3,000 abairt roghnaithe go randamach ó chorpas NCI (forbairt ar chlibeáil de réir na rann cainte).	Clibeáil uathoibríoch bunaithe ar rialacha, le fíorú agus ceartú láimhe.
3. Bunachar Crainn Spleáchas Gaeilge (BCSG) *https://github.com/tlynn747/IrishDependencyTreebank* Bunachar Crainn Spleáchas Uilíoch Gaeilge (BCSUG) *http://universaldependencies.org/#ga*	1,020 abairt pharsáilte de theanga scríofa (caighdeán óir). 4,910 abairt pharsáilte de theanga scríofa (caighdeán óir). Forbairt idir lámha.	Bunaithe ar scéim anótála faoi anáil LFG. Parsáilte go huathoibríoch, agus seiceáil láimhe agus ceartú ina dhiaidh sin. Leagan UD: BCSG mapáilte chuig scéim anótála BCSUG.
4. Corpas Gréasáin Gaeilge Crúbadán *http://crubadan.org/languages/ga*	100+ milliún focal de shonraí scríofa ó ransú gréasáin – nuashonraithe go laethúil.	Ransaithe go huathoibríoch ar líne agus ailínithe go huathoibríoch.
5. Vicipéid *https://ga.wikipedia.org*	57k alt – Vicipéid Ghaeilge (ciclipéid ar líne).	Sonraí nasctha struchtúrtha (téacs agus íomhánna). Cruthaithe de láimh ag an bpobal - athbhreithnithe ag eagarthóirí Vicipéid.
6. Corpas Gaeilge Comhaimseartha Gaois *http://www.gaois.ie/g3m/ga/*	24.7 milliún focal de théacs comhaimseartha foilsithe ó thús an 21ú céad ar aghaidh.	Coimeádaíocht láimhe agus leath-uathoibríoch.
7. Corpas Comhthreomhar Gaois *https://www.gaois.ie/crp/ga/*	Corpas comhthreomhar de mhíreanna téacs ailínithe ó reachtaíocht na hÉireann agus an Aontais Eorpaigh. 26 milliún focal Gaeilge agus 24.5 milliún focal Béarla.	Coimeádaíocht láimhe agus leath-uathoibríoch.
8. Corpais de Théacs Meán Sóisialta *https://github.com/tlynn747/IrishTwitterPOS* *https://github.com/UniversalDependencies/UD_Irish-TwittIrish*	1,500 tvuít, iad leimithe agus clibeáilte de réir na roinne cainte. Bunachar Crainn Spleáchas Uilíoch Gaeilge Twitter ina bhfuil os cionn 800 tvuít as Gaeilge.	Réamhchlibeáilte go huathoibríoch le gnáthchlibeálaí de réir na roinne cainte. Mapáilte chuig tacar clibeanna roinne cainte Twitter agus fíoraithe de láimh. Mapáilte chuig tacar clibeanna roinne cainte UD: parsáilte trí BCSUG a bhrógstrapáil agus fíoraithe de láimh.
9. Corpas de theilgin chainte ilfhoclacha as Gaeilge	1,700 abairt as Gaeilge agus iad clibeáilte le teilgin chainte ilfhoclacha.	Anótáilte de láimh agus athbhreithniú iomlán déanta orthu de réir treoirlínte PARSEME.
10. EduGA: Corpas d'Ábhair Oideachais	7.5 milliún focal ó ábhair oideachais do mhúineadh na Gaeilge agus ábhair eile a mhúintear trí Ghaeilge.	Clibeáil uathoibríoch de réir na roinne cainte bunaithe ar rialacha le ceartúchán láimhe. Téacschomharthú uathoibríoch abairte le ceartú láimhe.
11. Corpas Foghlaimeoirí Gaeilge TEG (Corpas Foghlaimeoirí TEG)	Timpeall 200,000 focal de theanga scríofa agus labhartha ginte ag foghlaimeoirí ó thrialacha inniúlachta Theastas Eorpach na Gaeilge (TEG) ag an meánleibhéal go dtí an t-ardleibhéal.	Ceartú earráidí/marcáil de láimh agus clibeáil uathoibríoch roinne cainte bunaithe ar rialacha saincheaptha ina dhiaidh sin.
12. Corpas Comhthreomhar Béarla-Gaeilge an Aontais Eorpaigh *http://www.lr-coordination.eu/resources* Comhordú Eorpach Acmhainní Teanga (ELRC) *http://www.lr-coordination.eu/resources* eSTÓR *https://estor.ie*	Bailiú leanúnach sonraí aistriúcháin náisiúnta ag DCU trí bhíthin ELRC-SHARE/eSTÓR. Bailithe ó chomhlachtaí poiblí éagsúla agus ó ranna rialtais faoi chomhaontuithe éagsúla ceadúnaithe – Sonraí Oscailte (PSI) den chuid is mó. 195,000+ abairt chomhthreomhar. Ná sonraí á mbailiú ar mhaithe le meaisínaistriúchán a fhorbairt ar an leibhéal náisiúnta agus Eorpach. Fásfaidh sé le hionchur ó eSTÓR.	Téacsanna aistrithe Béarla agus Gaeilge ginte ag aistritheoirí gairmiúla, ailínithe de láimh agus go leath-uathoibríoch ar leibhéal na habairte.
13. Cnuasach Bhéaloideas Éireann – Bailiúchán na Scol (Gaeilge) *http://Duchas.ie*	Breis agus 3.8 milliún focal Gaeilge ó na 1930idí.	Tras-scríofa trí bhíthin sluafhoinsiú.
14. Corpas Stairiúil na Gaeilge 1600-1926 *http://corpas.ria.ie* *https://www.ria.ie/research-projects/focloir-stairiuil-na-gaeilge*	90 milliún focal i mbreis agus 3000 téacs a foilsíodh i nGaeilge idir 1600 agus 1926. Tá an corpas seo mar bhunús ag Foclóir Stairiúil Gaeilge Acadamh Ríoga na hÉireann.	Caighdeánú litrithe láimhe agus uathoibríoch agus clibeáil uathoibríoch roinne cainte ina dhiaidh sin bunaithe ar rialacha saincheaptha.
15. Corpas d'fhilíocht shiollach na Gaeilge, circa 1200-1650 *https://app.sketchengine.eu/#open*	1+ mhilliún focal de Nua-Ghaeilge Chlasaiceach, a bhfuil 300K focal acu ar fáil saor in aisce ar líne.	Caighdeánú litrithe de láimh agus uathoibríoch agus clibeáil uathoibríoch roinne cainte ina dhiaidh sin bunaithe ar rialacha saincheaptha.

Bhí corpais téacs i gceist le formhór na dtionscnamh corpais ar mhórscála go dtí seo (Tábla 3.1). Tá tús curtha le bailiú corpas urlabhra don teicneolaíocht urlabhra le blianta beaga anuas. Tá roinnt samplaí de chorpas cainte agus labhartha san áireamh i dTábla 3.2.

Tábla 3.2: Corpais Urlabhra agus Labhartha

Corpus	Description	Annotation Details
1. Corpas Sintéise Ginearálta Urlabhra ABAIR: www.abair.ie	Idir lámha: 3 phríomhchanúint faoi láthair, c. 25 uair an chloig de chaint taifeadta.	Ranna deighilte agus anótáil foghraíochta ailínithe (X-SAMPA, IPA); marcáil aicinn.
2. Sintéis Urlabhra Chomhdhlúite ABAIR An Corpas Beag	Idir lámha: (c.2K leid) leideanna scripte le haghaidh clúdach foghraíochta uasta le méid íosta ábhair. Deartha faoi láthair do Ghaeilge na Mumhan. Obair idir lámha ar chanúintí eile .	Anótáil foghraíochta deighilte agus ailínithe (X-SAMPA, IPA); marcáil aicinn.
3. Corpas Labhartha Comhrá https://www.scss.tcd.ie/~uidhonne/comhra/	240K focal de theanga labhartha thras-scríofa spontáineach ó na mórchanúintí uile.	Athscríbhinní deighilte agus ailínithe le comhaid fuaime. Clibeáil uathoibríoch riailbhunaithe de réir na roinne cainte.
4. An Scéalaí: corpas foghlaimeoirí le haghaidh forbairt teicneolaíochta cainte	Á thógáil: sonraí á mbailiú ó ranganna sinsearacha bunscoile, meánscoile, treasach chomh maith le foghlaimeoirí fásta.	Gan anótáil fós.
5. Corpas Aitheanta Cainte ABAIR MíleGlór	Á thógáil: bailiú ar siúl trí bhíthin sluafhoinsiú agus taifeadtaí allamuigh i láithreacha Gaeltachta.	Tá eagarthóireacht agus mionchoigeartuithe ar an gcorpas idir lámha.
6. Corpas Foghlaimeoirí Gaeilge TEG (Corpas Foghlaimeoirí TEG)	Timpeall 200,000 focal de theanga scríofa agus labhartha ginte ag foghlaimeoirí tógtha as trialacha inniúlachta Theastas Eorpach na Gaeilge (TEG) ag an meánleibhéal go dtí an t-ardleibhéal.	Ceartú earráidí/marcáil de láimh agus clibeáil uathoibríoch roinne cainte ina dhiaidh sin bunaithe ar rialacha saincheaptha.
7. Corpas Inchomparáide Idirnáisiúnta ICCI – Gaeilge Labhartha agus Scríofa	Á thógáil. Sprioc: 600k focal de theanga labhartha spontáineach agus 400k focal de theanga scríofa.	Teastaíonn taifeadtaí nua go príomha ó shonraí labhartha. Sonraí scríofa á bhfoinsiú.

3.7 Corpais – Moltaí

3.7.1 Bailiú sonraí, cothabháil agus cruthú corpas idir lámha

I gcomhthéacs an chur chuige atá bunaithe ar shonraí agus ar staitisticí an lae inniu, is féidir, de thoradh corpais atá níos mó agus ar chaighdeán níos airde, uirlisí agus feidhmchláir teicneolaíochta teanga níos fearr a fhorbairt. Mar sin, cé go dteastaíonn corpais fhairsinge nua don Ghaeilge, ní mór na corpais atá liostaithe thuas a leathnú agus a choimeád ar bhonn leanúnach. Is gá pleanáil cháiréiseach a dhéanamh ar gach gné d'fhorbairt corpais amach anseo, is é sin ó thaobh líon agus méid na samplaí, na cainteoirí a ndírítear orthu agus a gcúlra teanga, an tréimhse ama atá i gceist, cineálacha teanga, cineálacha agus leibhéal anótála, srl, agus dá réir cinntiú go bhfuil na corpais oiriúnach don taighde sonrach agus don fhorbairt teicneolaíochta atá beartaithe. Ba cheart go mbeadh an cur chuige maidir le forbairt na gcorpas uile amach anseo cothrom le dáta leis na cineálacha cur chuige, na caighdeáin agus na cleachtais idirnáisiúnta is fearr dá bhfuil ann.

Seo a leanas roinnt moltaí sonracha:

- Tá corpais anótáilte ar chaighdeán óir (arna bhfíorú de láimh) riachtanach chun teicneolaíochtaí cainte agus teanga ar ardchaighdeán a fhorbairt agus is acmhainn luachmhar iad le haghaidh taighde teangeolaíoch. Mar shampla, tá gá le fairsingiú suntasach ar chorpais agus ar bhunachair chrainn ar a bhfuil clibeáil ar chaighdeán óir déanta de réir na roinne cainte ionas go mbeifear in ann parsálaithe láidre iontaofa a fhorbairt.
- Beidh gá le corpais chainte ar mhórscála a dhearadh, a bhailiú agus a phróiseáil chun freastal ar riachtanais na mbunteicneolaíochtaí: sintéis, córais aitheanta agus comhráiteacha cainte (Cuid II). Caithfidh siad seo freastal ar na canúintí éagsúla Gaeilge. Mar a pléadh i gCaibidlí 8, 9 agus 10, cé go mbeidh clúdach leathan (aois, inscne, srl.) i gceist le go leor corpas, beidh gá freisin le corpais shonracha (mar shampla glórtha leanaí) le haghaidh feidhmchláir spriocdhírithe.
- Corpais anótáilte go séimeantach. Toisc nach bhfuil corpas anótáilte séimeantach ar fáil don Ghaeilge faoi láthair, is tosaíocht é acmhainn dá leithéid a fhorbairt d'fhonn uirlisí agus feidhmchláir a bhfuil bonn eolais séimeantach fúthu a sholáthar, mar shampla foghlaim ríomhchuidithe teangacha, aisghabháil faisnéise, achoimre doiciméad agus parsáil láidir. Toisc gur iarracht mhionsonraithe í an anótáil shéimeantach a thógann am, d'fhéadfaí tús áite maidir le hanótáil a thabhairt do bhriathra ardmhinicíochta mar a rinneadh don Bhascais [3]. Chun barr éifeachtúlachta a bhaint amach, ba chóir na hacmhainní foclóireachta uile atá ar fáil (féach Caibidil 4) agus modhanna anótála leath-uathoibríocha (féach Caibidil 5) a úsáid chun an fhorbairt a bhrostú.
- Corpais a bhaineann go sonrach le fearann. De réir mar a mhéadaíonn an t-éileamh ar uirlisí saincheaptha próiseála teanga nádúrtha, feicfimid an t-éileamh ar chorpais a bhaineann go sonrach le fearann (mar shampla corpais alt eolaíochta, ábhar oideachais,

téacsanna dlí agus téacsanna eile a bhaineann go sonrach le fearann áirithe).
- Ní mór corpais scríofa an NCI a leathnú go suntasach chun freastal ar riachtanais na foclóireachta corpasbhunaithe nua-aimseartha. Ní mór méid suntasach teanga labhartha thras-scríofa a chuimsiú ann freisin. Toisc go bhfuil fearainn mheasctha san NCI, tá meiteashonraí mionsonraithe tábhachtach chun abairtí a roghnú nó a scagadh amach bunaithe ar sheánra nó ar chineál na foinse. Tá an cineál seo corpais mhóir ríthábhachtach freisin chun leabuithe focal a oiliúint le haghaidh uirlisí PTN, samhlacha teanga le haghaidh meaisínaistriúchán staitistiúil nó tuartheacs agus ais-aistriúchán chun sonraí sintéiseacha a chruthú le haghaidh meaisínaistriúchán néarach (Caibidil 5).
- Teastaíonn roinnt cineálacha corpas labhartha chun tacú le teicneolaíocht urlabhra agus teanga, lena n-áirítear corpais labhartha a bhaineann go sonrach le fearann, rud atá riachtanach chun sainghnéithe foclóra agus dioscúrsa a shainaithint agus a chainníochtú.
- Ní mór an Vicipéid a leathnú go leanúnach mar fhoinse leanúnach de théacsanna digiteacha atá ar fáil saor in aisce. Toisc gur tionscnamh faoi thionchar an phobail í an Vicipéid, tá iarracht ar fud na tíre ag teastáil chun úsáid na Vicipéide a chur chun cinn, mar aon le hoiliúint a chur ar eagarthóirí deonacha nua. Le cineálacha cur chuige nuálaíocha a mbíonn institiúidí oideachais páirteach iontu, beifear in ann fás leanúnach na hacmhainne seo a chinntiú.
- Beidh gá le hábhar a ghintear ag úsáideoirí (mar shampla sonraí idirlín agus meán sóisialta) a bhailiú, a choimeád agus a anailísiú chun léargas cothrom le dáta a choinneáil ar an gcaoi a n-úsáidtear an teanga agus an chaoi a bhforbraíonn sí ar líne. Braitheann uirlisí anailíse seintiminte (measúnú ar thuairimí a roinntear ar líne nó i gcáipéisí mórscála) freisin ar chorpais den chineál sin a bheith ar fáil.
- Tá cineálacha éagsúla corpas foghlaimeoirí (scríofa agus labhartha) ríthábhachtach chun uirlisí foghlama teanga a dhearadh faoi thionchar eimpíreach.

3.7.2 Uirlisí anótála, léirshamhlaithe agus fíoraithe corpas

Is tasc deacair é anótáil agus rangú téacs agus cainte a thógann cuid mhór ama agus éilíonn sé taighdeoirí ardoilte. Is éard atá i gceist leis ná anailís, catagóiriú agus lipéadú dlúth ar fhóinéimí, ar fhocail, ar nathanna cainte, ar ghaolmhaireachtaí idir focail, idir comhábhair in abairt nó idir tagróirí agus tagraithe i ndioscúrsa. I gcorpais chainte d'fhéadfadh go mbeadh gá freisin le sainaithint agus lipéadú a dhéanamh ar ghnéithe sainiúla fuaimiúla/urlabhraíochta (mar shampla buaicphointí toinne). Chun corpais mhóra ardchaighdeáin a tháirgeadh, moltar uirlisí anótála, léirshamhlaithe agus fíoraithe a fhorbairt a d'fhéadfadh na nótaí seo a dhéanamh indéanta ar scála mór, agus ar an gcaoi sin éascú a dhéanamh ar thaighde anailíseach. Ba chóir go gceadódh uirlisí den chineál sin freisin lipéadú éasca agus nascadh/dínascadh eilimintí.

Déantar cur síos i gCaibidil 13 ar ar an gcaoi ar féidir le sluafhoinsiú a bheith éifeachtach chun cineálacha áirithe fo-thascanna nach dteastaíonn sainscileanna lena n-aghaidh a scálú. I réimse leathan an PTN, úsáidtear ardáin sluafhoinsithe ar nós Amazon Mechanical Turk agus Figure Eight (Crowdflower roimhe seo) go forleathan mar mhodhanna chun an pobal a chur i dteagmháil le hanótáil sonraí. Mar sin féin, teastaíonn ardáin náisiúnta atá deartha go sonrach do chainteoirí Gaeilge chun an cur chuige seo a éascú lena chinntiú go bhfaightear na scileanna Gaeilge cuí.

3.7.3 Ardáin fiosraithe agus anailíse corpas

Chomh maith le corpais a fhorbairt, tá sé tábhachtach gur féidir na sonraí anótáilte sna hacmhainní seo a rochtain, a chuardach agus a anailísiú go héasca. Mar sin, is léir go bhfuil gá le hardáin oiriúnacha ar féidir le húsáideoirí (de gach cumas teicniúil) úsáid a bhaint astu chun na hacmhainní a bhrabhsáil nó a fhiosrú (mar shampla na corpais a bhfuil clibeanna na ranna cainte orthu, bunachair chrainn, corpais anótáilte foghraíochta, tras-scríbhinní, comhaid fuaime, srl.). Ba chóir go gceadódh na hardáin seo go ndéanfaí comhtháthú le huirlisí próiseála agus anailíse atá ann cheana. Moltar (i) iniúchadh a dhéanamh ar oiriúnacht na n-uirlisí foinse oscailte atá ann faoi láthair agus (ii) uirlisí nua atá sainoiriúnaithe don teanga a fhorbairt nuair is gá.

Tagairtí

1. Kilgarriff, A., M. Rundell, agus E. Uí Dhonnchadha, Efficient corpus development for lexicography: building the New Corpus for Ireland. Language Resources and Evaluation Journal, 2007. 40(2): lgh. 127-152.

2. Uí Dhonnchadha, E. agus J. van Genabith, Scaling an Irish FST morphology engine for use on unrestricted text, in Finite-State Methods in Natural Language Processing. 5th International Workshop, FSMNLP 2005, Helsinki, An Fhionlainn, 2005. Revised Papers., A. Yli-Jyr\"a, L. Karttunen, agus J. Karhum\"aki, Eagarthóirí. 2006, Springer International Publishing: Beirlín. lgh. 247–258.

3. Uí Dhonnchadha, E., A. Frenda, agus B. Vaughan, Issues in Designing a Corpus of Spoken Irish, in Proceedings of the 8th International Conference on Language Resources and Evaluation. 2012: Istanbul. lgh. 23–25.

4. Lynn, T., K. Scannell, and E. Maguire, Minority Language Twitter: Part-of-Speech Tagging and Analysis of Irish Tweets, in Proceedings of the ACL 2015 Workshop on Noisy User-generated Text, W. Xu, B. Han, agus A. Ritter, Eagarthóirí. 2015: Beising, An

tSín. lgh. 1--8.

5. Uí Dhonnchadha, E., Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar. 2009, Ollscoil Chathair Bhaile Átha Cliath.

6. Lynn, T., Irish Dependency Treebanking and Parsing. 2016, Ollscoil Chathair Bhaile Átha Cliath agus Ollscoil Macquarie. lch. 174.

7. Lynn, T. agus J. Foster, Universal Dependencies for Irish, in Proceedings of the Second Celtic Language Technology Workshop. 2016: Páras, An Fhrainc. p. 79--92.

8. Marcus, M., B. Santorini, agus M. Marcinkiewicz, Building a Large Annotated Corpus of English: The Penn Treebank. Computational Linguistics, 1993. 19(2): p. 313.

9. Futrell, R., K. Mahowald, agus E. Gibson, Quantifying Word Order Freedom in Dependency Corpora, in Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015). 2015: Uppsala, An tSualainn. lgh. 91–100.

10. Duong, L., et al., A Neural Network Model for Low-resource Universal Dependency Parsing, in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), L.i. M\`arquez, C. Callison-Burch, agus J. Su, Eagarthóirí. 2015, Association for Computational Linguistics: Liospóin, An Phortaingéil. lgh. 339–348.

11. Zeman, D., et al., CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, in Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. 2018, Association for Computational Linguistics: An Bhruiséil, An Bheilg. lgh. 1–21.

12. Straka, M. agus J. Straková, Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe, in Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. 2017, Association for Computational Linguistics: Vancouver, Ceanada. lgh. 88-99.

13. Baldwin, T. agus SN Kim, Multiword expressions, in Handbook of Natural Language Processing, An Dara hEagrán, N. Indurkhya agus F.J. Damerau, Eagarthóirí. 2010, CRC Press: Boca Raton, SAM. lgh. 1–39.

14. Monti, J. agus T. Amalia, Multiword Units Translation Evaluation in Machine Translation: Another Pain in the Neck?, in Proceedings of Multi-word Units in Machine Translation and Translation Technologies (MUMTTT 2015), M\'alaga, An Spáinn, G. Corpas Pastor, et al., Eagarthóirí. 2015, Eagráin Tradulex: M\'alaga, An Spáinn. lgh. 25–30.

15. Walsh, A., T. Lynn, agus J. Foster, Ilfhocail: A Lexicon of Irish MWEs, in Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019), A. Savary, et al., Eagarthóirí. 2019: Firenze, An Iodáil. lgh. 162–168.

16. Scannell, K. An Crúbadán. Ar fáil ó: http://crubadan.org/.

17. Scannell, K., Statistical models for text normalization and machine translation, in Proceedings of the First Celtic Language Technology Workshop, A Workshop of the 25th International Conference on Computational Linguistics (COLING 2014), 23 Lúnasa 2014, J. Judge, et al., Eagarthóirí. 2014: Baile Átha Cliath, Éire. lgh. 33--40.

18. Scannell, K., Standardization of corpus texts for the New English-Irish Dictionary, in NAACLT conference. 2009, http://borel.slu.edu/pub/naaclt09.pdf: Nua-Eabhrac.

19. Cassidy, L., T. Lynn, J. Barry, and J. Foster. TwittIrish: A Universal Dependencies Treebank of Tweets in Modern Irish. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), lgh 6869–6884, Baile Átha Cliath, Éire. Association for Computational Linguistics.

20. Lynn, T. agus K. Scannell, Code-switching in Irish tweets : A preliminary analysis, in Proceedings of the Celtic Language Technology Workshop, T. Lynn, et al., Eagarthóirí. 2019, European Association for Machine Translation: Baile Átha Cliath, Éire. p. 32--40.

21. Cortes S, et al., Brazilator: Machine Translation & Sentiment Analysis for World Cup 2014. Baile Átha Cliath, Éire; 2014. 2014.

22. Afli, H., S. McGuire, agus A. Way. Sentiment translation for low resourced languages: Experiments on Irish general election tweets. in 18th International Conference on Computational Linguistics and Intelligent Text Processing. 2017. Búdaipeist, An Ungáir.

23. An Phríomh-Oifig Staidrimh, Daonáireamh 2016. 2016, An Phríomh-Oifig Staidrimh: Baile Átha Cliath.

24. Ó Meachair, M.J., The Creation and Complexity Analysis of a Corpus of Educational Materials in Irish (EduGA). 2019, Coláiste na Tríonóide, Baile Átha Cliath. lch. 208.

25. Ní Ghloinn, A., E. Uí Dhonnchadha, agus A. O'Keeffe. The design and annotation of the TEG learner corpus of Irish. in Inter-Varietal Applied Corpus Studies

International Biennial Conference. 2018. Málta.

26. Dowling, M., et al., Tapadóir: Developing a Statistical Machine Translation Engine and Associated Resources for Irish, in The 4th LRL Workshop: Language Technologies in support of Less-Resourced Languages, G. Pavlovi\'c La\vzeti\'c, et al., Eagarthóirí. 2015: Poznan, An Pholainn.

27. Papavassiliou, V., P. Prokopidis, agus G. Thurmair, A modular open-source focused crawler for mining monolingual and bilingual corpora from the web, in Proceedings of the Sixth Workshop on Building and Using Comparable Corpora, S. Sharoff, R. Rapp, agus P. Zweigenbaum, Eagarthóirí. 2013, Association for Computational Linguistics: Sóifia, An Bhulgáir. lgh. 43–51.

28. Esplà, M., et al., ParaCrawl: Web-scale parallel corpora for the languages of the EU, in Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks. 2019, European Association for Machine Translation: Baile Átha Cliath, Éire. lgh. 118–119.

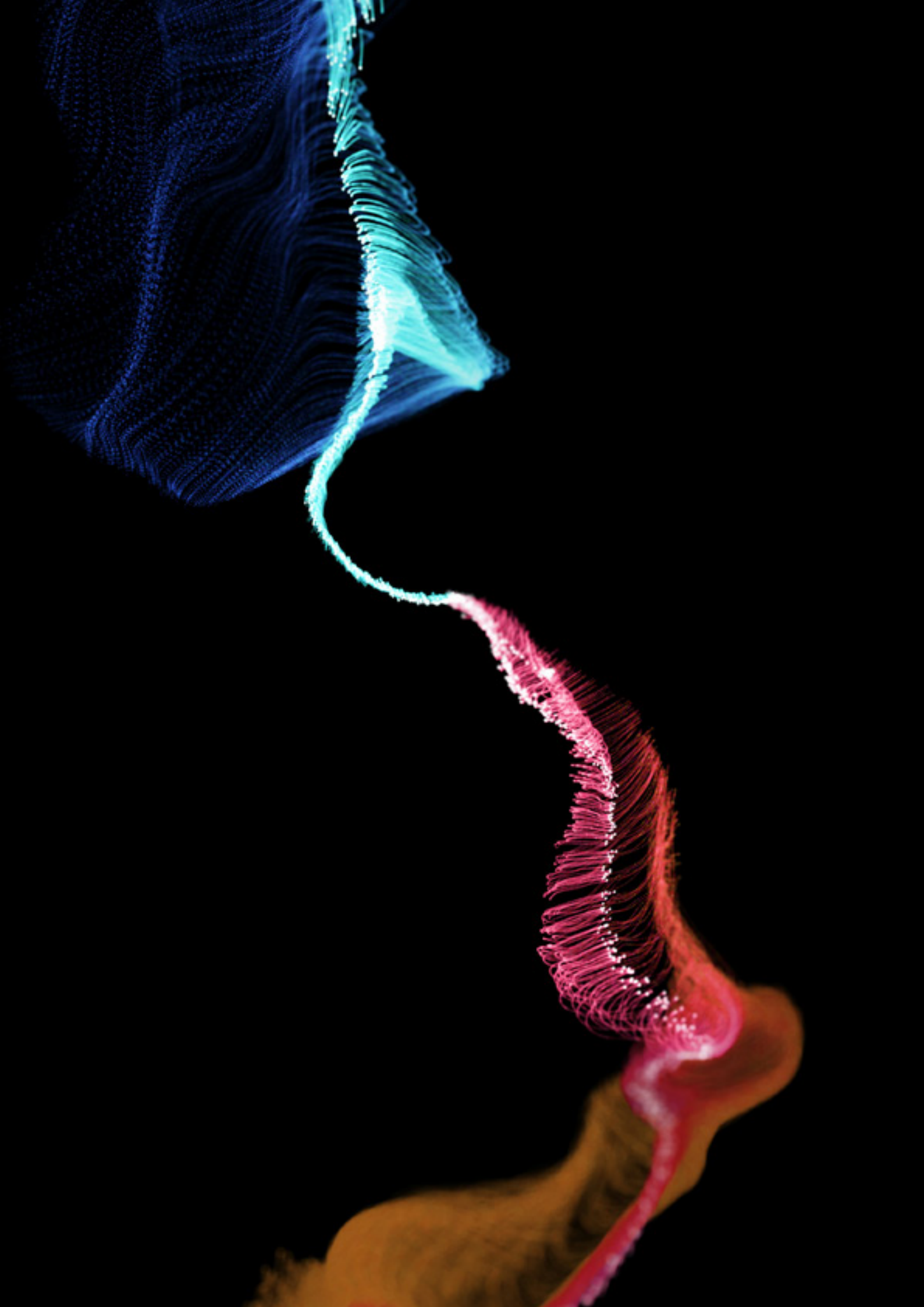
29. Ó Raghallaigh, B., K. Scannell, agus M. Dowling, Improving full-text search results on dúchas.ie using language technology. , in Celtic Language Technology Workshop. 2019: Baile Átha Cliath. lgh. 63–9.

30. Uí Dhonnchadha, E., et al. Corpas na Gaeilge 1882-1926: Integrating Historical and Modern Irish Texts. in LREC 2014 Workshop LRT4HDA: Language Resources and Technologies for Processing and Linking Historical Documents and Archives. 2014. Reykjavik, An Íoslainn.

31. Wigger, A., Caint Chonamara: Bailiúchán Hartmann. 2000.

32. Kirk, J. agus A. Čermáková. From ICE to ICC: The new International Comparable Corpus. in Workshop on Challenges in the Management of Large Corpora and Big Data and Natural Language Processing (CMLC-5+BigNLP) 2017. 2017. Birmingham.

33. Sharoff, S., et al., Eagarthóirí. BUCC - Building and Using Comparable Corpora. 2013, Springer.



Caibidil 4
Bunachair Faisnéise Teangeolaíochta

4.1 Céard is Bunachar Faisnéise Teangeolaíochta ann?

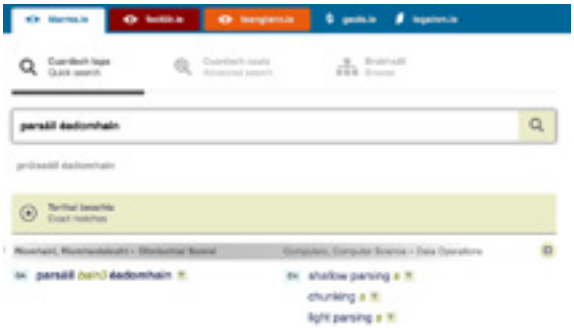
Is éard is bunachair faisnéise teangeolaíochta ann ná bailiúcháin eagraithe de shonraí teanga. Samplaí d'acmhainní teanga den chineál sin is ea foclóirí leictreonacha, bunachair sonraí théarmaíochta, teasárais, gasaitéir (daoine agus áiteanna) agus gluaiseanna. I mbunachair faisnéise coinnítear réimse faisnéise – faisnéis aistriúcháin agus bhrí, faisnéis ghramadúil, agus uaireanta sonraí úsáide i gcomhthéacs. Is féidir faisnéis theangeolaíoch a bheith iontu faoi ghnéithe éagsúla den teanga, amhail focail aonair, aonáin ilfhoclacha agus na naisc bhríocha eatarthu. Cé go bhfuil abairtí agus míreanna cainte a tharlaíonn go nádúrtha i gcorpais (Caibidil 3), féadfaidh sé nach mbeidh gach foirm fhéideartha den fhocal iontu. Ar an lámh eile, féachann bunachair faisnéise teangeolaíochta le fardail/liostaí d'fhoirmeacha focail a choinneáil agus an chaoi lenar féidir iad a úsáid. Ar an mbealach seo, cuireann corpais agus bunachair faisnéise faisnéis chomhlántach ar fáil faoi theanga; cuireann corpais faisnéis faoi ghnáthúsáid agus na foirmeacha is minice ar fáil; agus cuireann bunachair faisnéise faisnéis faoi úsáidí agus srianta féideartha ar fáil.

Sachaibidilseo,déanaimidcursíosarthríphríomhchatagóir de bhunachair faisnéise: bunachair faisnéise léacsaí, comhréire agus séimeantaice. I mbunachair faisnéise léacsaí coinnítear faisnéis ag leibhéal an fhocail agus an leibhéal ilfhoclach. I mbunachair faisnéise comhréire coinnítear faisnéis faoin gcaoi ar féidir focail a chur le chéile chun abairtí gramadúla a dhéanamh. I mbunachair faisnéise séimeantaí cuimsítear faisnéis maidir le bríonna focail aonair nó teaglaim focal.

4.2 Cén fáth a bhfuil Bunachair Faisnéise Teangeolaíochta tábhachtach?

Mar atá i gceist le corpais, tá bunachair faisnéise riachtanach sa taighde teicneolaíochta teanga agus leagtar béim ar a luach i bhformhór chaibidlí na cáipéise seo. Ag an leibhéal an-bhunúsach, i gcás seiceáil litrithe, is féidir fíorú a dhéanamh ar fhocal a bheith i dteanga le ríomhaire trí bhunachar sonraí léacsacha a sheiceáil. Is féidir bunachar sonraí léacsacha a úsáid chomh maith in uirlis idirghníomhach a ligeann d'úsáideoirí ar líne an luchóg a fhágáil os cionn focail nó cliceáil ar fhocal nach eol dóibh d'fhonn rochtain a fháil ar iontráil bhrí nó ar iontráil aistriúcháin i bhfoclóir (féach Caibidil 13).

Is acmhainní iad bunachair sonraí théarmaíochta dhátheangacha atá úsáideach go sonrach chun an téarmaíocht cheart a fháil le húsáid in aistriúcháin. Is minic a úsáideann foireann aistritheoirí na taisclanna comhroinnte seo chun comhsheasmhacht téarmaíochta a chinntiú i dtimpeallacht aistriúcháin a fhaigheann tacaíocht theicniúil (féach Caibidil 11), go háirithe i gcás fearainn ar leith, amhail na fearainn dlí, leighis, oideachais srl. I bhFíor 1 taispeántar cuardach ar iontráil don réimse teicneolaíochta teanga. Tá bunachair sonraí théarmaíochta luachmhar chomh maith chun uirlisí anailíse séimeantaí (Caibidil 5) agus uirlisí foghlama teanga a fhorbairt (féach Caibidil 14). Tá gá chomh maith le tuilleadh foclóirí aonteangacha don Ghaeilge. Foclóirí iad seo ina dtugtar tuairisc ar bhrí agus úsáid ghramadúil na bhfocal Gaeilge trí mheán na Gaeilge féin. Má tá úsáideoir foclóra ag seiceáil na hiontrála le haghaidh focal Gaeilge, agus é ag foghlaim nó ag obair trí Ghaeilge, níor chóir go mbeifí ag súil go n-iompódh sé nó sí ar an mBéarla agus rochtain á fáil ar acmhainn léacsach. Tá sé seo fíor go háirithe i gcás daltaí scoileanna lán-Ghaeilge.



Fíor 1: Cuardach téarmaíochta ar tearma.ie

D'fhonn seiceáil gramadaí a dhéanamh ar dhoiciméad, ní mór eolas a bheith ag ríomhairí faoi rialacha gramadaí teanga. Tá an t-eolas seo riachtanach chun teanga a thuiscint agus a ghiniúint go cruinn. Mar shampla, ní abairt ghramadúil an abairt *Thug an stiúrthóir doiciméad 'The director gave a document' mar gheall go bhfuil ceann de rannpháirtithe riachtanacha an bhriathair 'tabhair' ar iarraidh, is é sin an faighteoir.

D'fhéadfadh seiceálaí gramadaí riailbhunaithe rochtain a fháil ar fhaisnéis den chineál sin chun a fháil amach an sáraíonn abairt aon cheann de na srianta gramadaí. Is féidir eolas comhréire faoi ord focal, teaglaim focal agus caidrimh ghramadúla a ionchódú i ngraiméir ríomhaireachtúla. Chomh maith le seiceáil gramadaí (féach Caibidil 13), tá na cineálacha seo bunachair faisnéise bunúsach do mhórán tascanna próiseála teanga (féach Caibidil 5), agus feidhmeanna amhail giniúint teanga (féach Caibidil 6), asbhaint faisnéise (féach Caibidil 12), agus feidhmeanna foghlama teanga (féach Caibidil 14).

Ó thaobh brí de, tá eolas séimeantach tábhachtach do mhórán tascanna próiseála teanga. Mar shampla, i gcás inneall meaisínaistriúcháin riailbhunaithe a aistríonn ón nGaeilge go Béarla, teastaíonn bealach uaidh chun na bríonna éagsúla atá le fiach (`raven', 'debt', nó `hunt') a idirdhealú chun an t-aistriúchán ceart Béarla a sholáthar i gcomhthéacs. Is féidir feabhas a chur ar fheidhmíocht na n-inneall cuardaigh (Caibidil 12) trí bheith in ann débhríochtaí den chineál sin a réiteach, trí innéacsú doiciméad a chumasú de réir bríonna focail seachas díreach tarluithe de litriú focail.

Chomh maith le bunachair sonraí shéimeantacha aonteangacha, i gcomhthéacs ilteangach , m.sh. le haghaidh meaisínaistriúcháin, is gá bunachair sonraí shéimeantacha a nascadh idir teangacha. Mar gheall ar a dtábhacht sa Phróiseáil Teanga Nádúrtha (PTN), forbraíodh bunachair sonraí agus líonraí séimeantacha do líon mór teangacha, agus tá mórán acu nasctha le chéile. Saibhríonn sé seo bunachair sonraí théarmaíochta, cumasaíonn sé forbairt ar theasárais, agus cuidíonn sé le tascanna PTN amhail aithint aonáin ainmnithe (AAA) (Caibidil 5), meaisínaistriúchán (Caibidil 11), aisghabháil faisnéise trasteangaí (Caibidil 12), i measc nithe eile.

Is cineál bunachair faisnéise é DBpedia a cheadaíonn asbhaint níos fearr ar fhaisnéis ó ailt na Vicipéide. Trí DBpedia a bheith ar fáil do theanga, is féidir leathanaigh Vicipéide don teanga sin a chuardach ar bhealach níos cruinne agus níos éifeachtaí. Iarracht sluafhoinsithe pobail is ea DBpedia, amhail an Vicipéid. Tá DBpedia roinnte ina chaibidlí teanga. Tá gach caibidil teanga bunaithe ar obair dheonach (tionscnamh sluafhoinsithe) agus tá sí freagrach as faisnéis a asbhaint as leaganacha éagsúla teanga den Vicipéid.

Is cineál bunachar faisnéise é Sonraí Nasctha chomh maith. Is é aidhm na Sonraí Nasctha sonraí a léiriú i bhformáid struchtúrtha ionas gur féidir le ríomhairí na sonraí a léamh, a fhiosrú, a asbhaint agus a thuiscint. Ba chóir go soláthródh gach tacar sonraí de chuid Sonraí Nasctha naisc le tacair sonraí eile ionas gur féidir le clár ríomhaire nascleanúint a dhéanamh trí fhearainn éagsúla agus sonraí ábhartha breise a fháil. Toisc Sonraí Nasctha Oscailte (sonraí a scaoiltear faoi cheadúnas oscailte) a bheith ar fáil ar an ngréasán tá ar chumas forbróirí feidhmchlár na sonraí a úsáid láithreach.

4.3 Cén chaoi a n-oibríonn siad?

Tá ríomhairí ag brath go mór ar acmhainní teanga ina bhfuil faisnéis faoi na srianta agus na coinbhinsiúin léacsacha, comhréire agus séimeantaice a bhaineann le húsáid teanga. Sa chuid seo tugaimid léargas ginearálta ar ghnáthacmhainní léacsacha, comhréire agus séimeantaice.

Acmhainní léacsacha

Stóráiltear faisnéis faoi na focail aonair i dteanga i mbunachair faisnéise léacsacha. Féadfaidh siad a bheith i bhfoirm bunachair sonraí, líonraí agus léirithe digiteacha struchtúrtha eile sonraí amhail comhaid XML. Is minic a dhíorthaítear bunachair sonraí léacsacha ar dtús ó leaganacha meaisín-inléite d'fhoclóirí foilsithe, mar shampla [1-3] mar bhunús le suíomh gréasáin www.teanglann.ie.

Is gnách go mbíonn ceannfhocail in iontrálacha bunachar sonraí foclóra agus téarmaíochta, m.sh. bord seachas gach foirm lena n-áirítear bhord, bhoird, mbord, srl. Is féidir le huirlisí próiseála focal (féach Caibidil 13) foirmeacha focail a chuardach i mbunachar léacsach. Ó thaobh éifeachtúlacht ríomhaireachta de, áfach, is minic is féidir le feidhmchláir PTN leas a bhaint as uirlisí moirfeolaíochta (féach Caibidil 5) a dhéanann cuardach trí líonraí níos lú ina bhfuil leamaí nó ceannfhocail amháin, agus na rialacha chun na foirmeacha focail a bhaineann le ceannfhocal a ghiniúint.

I gcodarsnacht leis na foclóirí, a mbíonn claonadh iontu díriú ar theanga ghinearálta ilchuspóireach agus ar fhocail aonair, dírítear i mbunachair sonraí théarmaíochta ar theanga fhearannsonrach. Is minic a sholáthraíonn na bunachair sonraí seo samplaí de théarmaí in úsáid, agus an comhthéacs is oiriúnaí dóibh.

Ós rud é go ndéantar go leor faisnéise agus téarmaíochta a ainmniú go minic i mBéarla ar dtús (agus ansin iad a aistriú go teangacha eile), straitéis amháin a d'fhéadfadh a bheith ann chun bunachair théarmaíochta na Gaeilge a choinneáil cothrom le dáta is ea téarmaí nua a asbhaint agus a aithint go haonteangach i dtéacsanna teicniúla / speisialaithe Béarla nuafhoilsithe agus ansin na téarmaí sin a aistriú de láimh go Gaeilge. Ar an gcaoi sin tá téarmaíocht na Gaeilge réidh le húsáid nuair is gá cáipéisí i nGaeilge a chruthú.

Ina theannta sin, is féidir uirlisí agus modhanna sainoiriúnaithe [4] a úsáid chun cabhrú le hasbhaint téarmaíochta ó chorpais eolaíochta agus asbhaint téarmaíochta dátheangaí ó chuimhní aistriúcháin.

Acmhainní comhréire

Maidir le teicneolaíochtaí teanga a phróiseálann teanga ar scála níos mó ná leibhéal focal nó frásaí, is é sin ar leibhéal na habairte nó na cáipéise, éilíonn siad eolas ar rialacha comhréire na teanga. Tugtar parsáil

comhréire ar bhrath uathoibríoch struchtúr comhréire i dteanga (féach Caibidil 5). Ionchódaíonn graiméir ríomhaireachtúla léirithe ar struchtúr comhréire teanga agus is acmhainní riachtanacha iad do theicneolaíocht pharsála riailbhunaithe. I ngraiméir struchtúr frásaí agus i ngraiméir spleáchais coinnítear faisnéis faoi na cineálacha focal (ainmfhocail, briathra, aidiachtaí srl.) is féidir a úsáid le chéile agus na hoird a cheadaítear dóibh.

Is gnách gurb é briathar an chomhpháirt lárnach d'abairt agus is minic go mbíonn ceanglais agus srianta ag briathra maidir leis na focail áirithe is féidir a úsáid leo (m.sh. ainmfhocail, réamhfhocail srl.) agus cén áit a bhfuil siad suite in abairt maidir leis an mbriathar. Mar shampla, *chuaigh isteach sa rang an dalta (ord mícheart focal), *theip liom sa scrúdú (réamhfhocal mícheart), agus mar sin de, is struchtúir iad atá mícheart ó thaobh ghramadach na Gaeilge de. Baineann graiméir ríomhaireachta úsáid as acmhainní teanga amhail foclóirí fiúsachta briathra ina bhfuil faisnéis faoi líon na rannpháirtithe a bhaineann le briathar, m.sh. de ghnáth bíonn trí rannpháirtí de dhíth ar an mbriathar 'tabhair'; an deontóir, an rud atá á thabhairt agus an faighteoir. Is féidir foclóir fiúsachta a úsáid chun corpas anótáilte a chruthú amhail PropBank , (féach Caibidil 3).

Ionchódaíonn VerbNet[5, 6] struchtúr argóinte briathair, is é sin cé mhéad rannpháirtí agus cén cineál a theastaíonn ó bhriathar áirithe. Taispeántar i bhFíor 2 sampla den chaoi a ndéantar argóintí don bhriathar Béarla 'cut' a ionchódú in VerbNet. Thabharfadh forbairt acmhainne den chineál VerbNet don Ghaeilge faisnéis ar bhealach atá oiriúnach don ríomhaire, rud atá ríthábhachtach chun críocha forbartha graiméar, parsála agus oideolaíochta. Chomhlánódh sé anailís mhoirfeolaíoch chomh maith agus dhéanfadh sé comhtháthú le hacmhainní foclóireachta agus séimeantacha eile don Ghaeilge.

Fíor 2: Léiriú ar struchtúr na hargóinte le haghaidh 'cut' sa VerbNet Béarla

Acmhainní Séimeantacha

Ceann de na fadhbanna bunúsacha a bhaineann le ríomhairí a fháil le tuiscint a bhaint as teanga an duine is ea gur féidir le focail a bheith débhríoch ó thaobh brí (m.sh. an focal Gaeilge fiach, rud a chiallaíonn 'a debt', 'a raven', nó 'hunting' sa Bhéarla). Soláthraíonn bunachair sonraí shéimeantacha `fardal bríonna' do theanga, agus soláthraíonn siad freisin na bealaí inar féidir le ríomhaire idirdhealú a dhéanamh idir bríonna focal i gcomhthéacs. Go ginearálta, ó thaobh na hintleachta saorga de, ionchódaíonn líonra séimeantach cuid den `eolas ar an bhfíorshaol' a theastaíonn ó ríomhairí chun téacsanna a thuiscint agus a phróiseáil ar bhealach neamhfhánach.

Fíor 3: WordNet ar Líne

Is féidir eolas séimeantach a ionchódú mar líonraí caidrimh idir focail agus coincheapa. Tá líonraí séimeantaice níos saibhre ná teasárais thraidisiúnta, rud nach ndéanann ach taifeadadh ar chomhchiallaigh (gar) agus uaireanta frithchiallaigh. De ghnáth tugtar gréasáin focal ar na líonraí seo, tar éis Wordnet Béarla Princeton [7] atá ina bhunachar sonraí mór léacsach d'ainmfhocail, briathra, aidiachtaí agus dobhriathra Béarla (féach Fíor 3).

Grúpáiltear bríonna i dtacair de chomhchiallaigh chognaíocha (synset) bunaithe ar chosúlacht brí, coincheap ar leith á léiriú ag gach ceann acu. Tá na tacair synset idirnasctha trí bhíthin caidrimh choincheapúla-shéimeantacha agus léacsacha. Féach Fíor 4 mar a bhfuil sampla den chaoi a léirítear an téarma 'word' in acmhainn léacsach Princeton WordNet.

Fíor 4: Léiriú in WordNet ar an téarma 'word'

Pléann FrameNet [8] le habairtí seachas le focail (is é sin i gcodarsnacht le WordNet agus VerbNet), agus tá faisnéis ann ní hamháin ar shéimeantaic léacsach ach ar shéimeantaic argóinte faisnéise. Tá úsáid FrameNet ina réimse taighde atá ag fás sa PTN, mar shampla san anailís ar nathanna ilfhoclacha [9] agus parsáil shéimeantach[10].

Graif eolais

I réimse an léirithe eolais agus na réasúnaíochta, is bunachar eolais é graf eolais ina n-úsáidtear samhail sonraí grafstruchtúrtha nó toipeolaíocht chun sonraí a chomhtháthú. Is minic a úsáidtear graif eolais chun tuairiscí idirnasctha ar aonáin a stóráil – rudaí, imeachtaí, staideanna nó coincheapa teibí – agus ionchódaítear iontu freisin an tséimeantaic atá mar bhunús leis an téarmaíocht a úsáidtear. Tá dlúthbhaint ag graif eolais leis an nGréasán Séimeantach agus úsáidtear go forleathan iad in innill chuardaigh amhail Google agus Bing, i líonraí sóisialta amhail LinkedIn agus Facebook, i gcórais mholtóra amhail Amazon agus i gcórais freagartha ceisteanna amhail Siri nó Alexa. Is réitigh chumhachtacha PTN iad a chomhcheanglaítear go minic le córais mheaisínfhoghlama i gcur chuige ar a dtugtar Intleacht Shaorga Shiombalach. Tagann an chumhacht atá iontu ó thacair sonraí fholuiteacha atá lipéadaithe de réir tacsanomaíochtaí agus ointeolaíochtaí agus a chabhraíonn le greim a fháil ar ghaoil bhríocha (séimeantach) idir aonáin (m.sh. táirgí, daoine, gnólachtaí srl.).

4.4 A bhfuil déanta i leith na Gaeilge go dtí seo

Sa chuid seo tugaimid léargas ginearálta ar chuid de na príomhacmhainní léacsacha, comhréire agus séimeantaice atá á bhforbairt don Ghaeilge.

Acmhainní léacsacha

Go dtí seo, tá maoiniú díreach leanúnach faighte i leith acmhainní léacsacha don Ghaeilge a fhorbairt (m.sh. trí Fhoras na Gaeilge) agus mar sin tá dul chun cinn seasta léirithe le roinnt blianta anuas. I measc na n-acmhainní reatha tá bunachair sonraí téarmaíochta, foclóirí dátheangacha agus foclóir aonteangach, agus tá siad seo ar fáil don phobal trí shuíomhanna gréasáin nó aipeanna fóin chliste (foclóir.ie & teanglann.ie). Ina theannta sin, tá bunachar sonraí téarmaíochta náisiúnta na Gaeilge (téarma.ie) inrochtana go saor ar líne agus tá sé in úsáid go forleathan ag an bpobal i gcoitinne, institiúidí oideachais, aistritheoirí sa riarachán poiblí, in institiúidí Eorpacha agus ag saoraistritheoirí. Tugtar achoimre i dTábla 4.1 ar na hacmhainní éagsúla léacsacha atá á bhforbairt faoi láthair don Ghaeilge.

Ainm na hAcmhainne	Cur síos	Sonraí
Foclóir.ie *www.foclóir.ie*	An Foclóir Nua Béarla-Gaeilge (Foras na Gaeilge)	• 50,000 (48,056) ceannfhocal • 125,000 (123,309) coincheap
Leabharlann Teanga agus Foclóireachta Teanglann *www.teanglann.ie*	(Foras na Gaeilge)	• Foclóir Gaeilge-Béarla (Ó Dónaill, 1977): 42,106 • Foclóir Béarla-Gaeilge (de Bhaldraithe, 1959): 43,082 ceannfhocal • An Foclóir Beag (aonteangach) (Ó Dónaill & Ua Maoileoin, 1991): 16,753 ceannfhocal • Bunachar Sonraí Morfeolaíochta: 43,000 iontráil
Teasáras *www.potafocal.com*	Teasáras Gaeilge bunaithe ar an WordNet Gaeilge (Líonra Séimeantach na Gaeilge)	• Teasáras: 30,000+ iontráil ginte go huathoibríoch
Gluais Tí Gluais Beo *www.potafocal.com*	Bunachar sonraí inchuardaithe de ghluaiseanna dátheangacha Tugtar comhthéacs úsáide, aistriúcháin agus faisnéis ar inscne, úsáid ghramadúil, na ranna cainte agus minicíocht úsáide	• Gluais Tí: 5,000 iontráil • Gluais Beo!: 200,000 sampla úsáide
Téarma.ie – an Bunachar Náisiúnta Téarmaíochta don Ghaeilge *www.téarma.ie*	An Coiste Téarmaíochta (Fiontar & Scoil na Gaeilge , OCBÁC)	• Breis agus 183,000 téarma Gaeilge
Bunachar Sonraí Téarmaíochta IATE *www.gaois.ie/trm*	Tionscadal Lex AE (an Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meán agus Fiontar, OCBÁC)	• Breis agus 130,916 téarma dlí Gaeilge • Bailiúchán nathanna cainte: 420 nath cainte ilfhoclach
Logainm.ie *www.logainm.ie*	Bunachar Sonraí Logainmneacha (Fiontar, OCBÁC agus an Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meán)	• 200,000 péire logainmneacha Gaeilge-Béarla • 100,000 logainm geo-chlibeáilte
www.Ainm.ie	Bunachar sonraí bunaithe ar an tsraith Beathaisnéis, ina bhfuil daoine a raibh tionchar suntasach acu ar an nGaeilge.	• 1,769 beathaisnéis choimeádaithe i bhformáid sonraí nasctha
ilFhocail	Bunachar sonraí de nathanna ilfhoclacha Gaeilge á dtógáil ag Ionad ADAPT, OCBÁC.	• Breis agus 200,000 nath cainte ilfhoclach Gaeilge (ainmniúil agus briathartha) bailithe ó acmhainní éagsúla.

Tábla 4.1: Léargas ginearálta ar na hacmhainní sonraí teanga atá ar fáil don Ghaeilge

Acmhainní Comhréire agus Séimeantaice

Sa chuid seo, tugaimid léargas ginearálta ar na hacmhainní comhréire agus séimeantaice atá á bhforbairt don Ghaeilge. Tá na hacmhainní seo i gcéimeanna luatha na forbartha den chuid is mó agus teastaíonn tacaíocht airgeadais shuntasach uathu chun iad a thabhairt chuig an leibhéal a theastaíonn le go mbeidh siad éifeachtach i bhfeidhmchláir theanga.

Is foclóir fiúsachta briathra don Ghaeilge é Foclóir Briathra Gaeilge. San fhoclóir sin tugtar sonraí struchtúrtha i dtaobh thart ar dhá chéad briathar Gaeilge, agus tugtar cineálacha, bríonna agus patrúin úsáide, agus comhlogaigh thipiciúla (focail a úsáidtear in éineacht leo). Taispeánantar i bhFíor 5 an fhaisnéis fiúsachta don bhriathar 'abair'. D'fhorbair an tOllamh Arndt Wigger an foclóir fiúsachta seo sa Ghearmáin , trí úsáid a bhaint as Caint Chonamara , corpas mór labhartha a bailíodh in 1964 [11], agus Corpas Náisiúnta na Gaeilge [12], corpas ocht milliún focal den Ghaeilge scríofa. Tá na sonraí, arb ionann iad agus thart ar 5,500 iontráil, ar fáil i bhformáid XML agus tá siad ar fáil ar an suíomh gréasáin Potafocal.ie (maoinithe ag an gComhairle um Oideachas Gaeltachta agus Gaelscolaíochta).

Fíor 5: Foclóir Briathra Gaeilge (Foclóir Fiúsachta Briathra Gaeilge)

Feidhmíonn graiméir ríomhaireachta mar bhunús le tuairiscí comhréire agus ríomhaireachtúla ar an nGaeilge. Soláthraíonn siad tuairiscí eolaíocha ar theanga ina gcuirtear patrúin agus rialacha teanga in iúl i bhfoirm atá intuigthe ó thaobh na ríomhaireachta, ionas gur féidir feidhmchláir próiseála teanga a dhearadh (féach Caibidil 5).

I dtaighde ríomhtheangeolaíochta le déanaí sa réimse seo rinneadh doiciméadú ar mhórán de na struchtúir agus na patrúin a tharlaíonn sa Ghaeilge [13-17], ach is gá, áfach, graiméar ríomhaireachtúil cuimsitheach don Ghaeilge a fhorbairt fós.

Líonra séimeantach mórscála don Ghaeilge is ea Líonra Séimeantach na Gaeilge (LSG)[18] a cruthaíodh trí mhapáil a dhéanamh ar bhríonna focail ó WordNet Princeton (Béarla) go Gaeilge, go leath-uathoibríoch (féach Fíor 6). Ciallaíonn sé seo go bhfuil sé 'nasctha' go huathoibríoch leis an WordNet Béarla, agus dá bhrí sin, tá sé nasctha le líonraí séimeantacha don iliomad teangacha eile. Cruthaíodh leagan le déanaí de LSG a rinne na naisc seo follasach mar chuid den taighde ar mheaisínaistriúchán Béarla-Gaeilge[19].

Toisc go raibh an LSG bunaithe ar an WordNet Béarla, beidh mapálacha séimeantacha ann nach bhfuil le fáil i nGaeilge, agus os a choinne sin d'fhéadfadh go mbeadh mapálacha in easnamh orthu a bhaineann leis an nGaeilge. Is gá tuilleadh fíoraithe de láimh ar na naisc shéimeantacha a ghintear go huathoibríoch idir an Béarla agus an Ghaeilge, chun a chinntiú go bhfuil an WordNet curtha in oiriúint don Ghaeilge. Tá tacaíocht leanúnach riachtanach chomh maith toisc go gcaithfear a gcaidreamh séimeantach le hiontrálacha atá ann cheana a shainaithint agus a lipéadú. Is féidir acmhainn fíoraithe den chineál sin a úsáid mar theasáras le haghaidh scríbhneoireachta i nGaeilge (Caibidil 13), agus i bhfeidhmeanna amhail aisghabháil faisnéise Gaeilge (Caibidil 12). D'fhéadfadh sé a bheith ina chomhpháirt riachtanach freisin d'uirlisí éagsúla PTN amhail córais imdhealaithe débhríochta bríonna clúdaigh leathain nó parsálaithe séimeantacha (Caibidil 5). Tá an LSG ina bhunachar sonraí foinse oscailte agus ar fáil saor in aisce le húsáid ag taighdeoirí agus forbróirí bogearraí PTN.

Fíor 6 – Léiriú LSG (Wordnet Gaeilge) ar an téarma 'focal'

Wordnet Ilteangach

In Wordnet Ilteangach baintear úsáid as gréasáin focal atá ann cheana ó Open Multilingual Wordnet agus ó DBpedia chun gréasáin focal nua a ghiniúint do theangacha breise. Baineadh úsáid as an gcur chuige seo chun gréasáin focal a ghiniúint do bhreis agus 20 teanga lena n-áirítear an Ghaeilge [20]. D'éileodh sé seo freisin fíorú de láimh a dhéanamh ar naisc a ghintear go huathoibríoch chun deimhin a dhéanamh de go bhfuil siad sách cruinn le haghaidh feidhmeanna iartheachtacha.

Is é DBpedia as Gaeilge an chaibidil atá freagrach as an leagan Gaeilge de DBpedia a chothabháil. Baineann an chaibidil seo sonraí ón Vicipéid (Wikipedia na Gaeilge), á gcur ar fáil lena bhfiosrú agus a nascadh le tacair sonraí eile. Aistríonn sé na téarmaí Béarla freisin ó ointeolaíocht DBpedia. Ciallaíonn sé sin gur féidir aon fheidhmchlár a chruthaítear ó ointeolaíocht den chineál sin a logánú go Gaeilge go héasca amach anseo.

Tá sonraí ó Logainm.ie (féach Tábla 4.1), atá ar fáil mar Shonraí Nasctha Oscailte, in úsáid anois ag Suirbhéireacht Ordanáis Éireann, Eircode, GeoDirectory, OpenStreetMap (Fíor 8), agus mórán eile ionas gur féidir fiosrú áitiúil Gaeilge a dhéanamh. Tá acmhainn mar seo tábhachtach do thascanna PTN mar Aithint Aonáin Ainmnithe (féach Caibidil 5) agus aon fheidhmchlár cliste a éilíonn sainaithint geoshuímh (Caibidil 13).

4.5 Acmhainní Teanga: Bunachair Faisnéise – Moltaí Ginearálta

Is acmhainní bunúsacha iad bunachair faisnéise i bhforbairt uirlisí agus feidhmchlár teanga. Cé go ndearnadh roinnt dul chun cinn san am atá thart le hacmhainní den chineál seo, tugtar achoimre anseo ar na riachtanais lena chinntiú go bhfuil bonn láidir faoin nGaeilge chun tuilleadh forbartha a dhéanamh ar na huirlisí agus na feidhmchláir riachtanacha a leagtar amach sa doiciméad seo – ar dtús le moladh ginearálta, agus ina dhiaidh sin moltaí de réir gach cineáil bunachair faisnéise:

- Teastaíonn óstáil agus cothabháil leanúnach ar chomhéadain úsáideora chun an toradh uasta ar an infheistíocht a chinntiú agus na hacmhainní luachmhara seo á gcruthú, agus chun deimhin a dhéanamh de go mbeidh fáil leanúnach ag an bpobal orthu i bhfad na haimsire.
- Teastaíonn cothabháil agus nuashonrú rialta ó bhonneagair bunachar sonraí d'fhonn coinneáil suas leis an teicneolaíocht is déanaí ¬¬– teicneolaíocht a bhíonn ag síorathrú.
- Ba chóir an fhorbairt ar bhunachar faisnéise nua a bheith feidhmchlár-threoraithe chun an ábharthacht agus an inúsáidteacht is fearr a chinntiú.
- Ní mór bunachair sonraí léacsacha agus bunachair sonraí téarmaíochta fearannsonracha a choinneáil cothrom le dáta le focail agus téarmaíocht nua-aitheanta agus doiciméadaithe sa Ghaeilge.
- De réir mar a thiocfaidh méadú ar líon na mbunachar faisnéise léacsach neamhspleách don Ghaeilge, beidh gá le comhordú agus comhroinnt sonraí níos mó chun comhsheasmhacht agus comhoiriúnacht a chinntiú. Mar shampla, athraíonn litriú líon beag focal idir foclóir.ie, téarma.ie, agus an Foclóir Gaeilge-Béarla, rud a d'fhéadfadh mearbhall a chruthú i measc úsáideoirí agus córas araon. Ba cheart, ar a laghad, do choimeádaithe na mbunachar sonraí seo comhordú a dhéanamh nuair a chuirtear athruithe ortagrafacha nó gramadaí i bhfeidhm; go hidéalach, ba chóir na bunachair sonraí go léir a nascadh le chéile ionas gur féidir seiceálacha comhsheasmhachta uathoibrithe a chur i gcrích.
- Is gá forbairt a dhéanamh ar shaoráid cuardaigh chomhtháite le húsáid ag an bpobal a dhéanann cuardach ar iliomad bunachar sonraí mar shampla Téarma.ie, Foclóir.ie agus Teanglann.ie.
- Ba chóir sluafhoinsiú a spreagadh nuair is iomchuí, mar atá amhlaidh i gcás sonraí logainmneacha na hÉireann is féidir leis an bpobal a chur faoi bhráid Mheitheal Logainm.ie (féach Caibidil 15).

Acmhainní Léacsacha – Forbairtí Riachtanacha Amach Anseo

Tá an méid seo i measc cuid de na riachtanais is práinní:

- Teastaíonn bunachair sonraí théarmaíochta nua fhearannsonracha laistigh de thimpeallacht aistriúcháin ríomhchuidithe, m.sh. sa réimse dlí, ábhar teagaisc eolaíochta.
- Is gá foclóir Gaeilge aonteangach cothrom le dáta a úsáid i scoileanna lán-Ghaeilge, agus in ionaid oibre lán-Ghaeilge.
- Bunachar sonraí idirghníomhach de nathanna ilfhoclacha Gaeilge (m.sh. leaganacha cainte, struchtúir éadroma briathra, briathra míre, ainmfhocail chumaisc) ina ndéantar cur síos ar úsáid na bhfocal i gcomhthéacs focail a thagann

aníos go máguaird go minic, mar shampla Ilfhocal [21].
- Tuilleadh forbartha ar bhunachar sonraí Logainm.ie, go príomha chun ainmneacha bailte fearainn agus sráidainmneacha a chur leis, agus a fheidhmiúlacht seirbhíse gréasáin a fheabhsú.
- Úsáid téarmamhéadracht chun úsáid (nó neamh-úsáid) téarmaí a chinneadh le himeacht ama chun cinntiú go bhfuil bunachair sonraí cothrom le dáta (mar shampla féinphic vs. féinín do 'selfie')
- Cothabháil leanúnach ar fhoclóirí agus ar bhunachair théarmaíochta trí théarmaí a asbhaint ó chorpais (féach Caibidil 3).
- Uirlisí agus modhanna atá sainoiriúnaithe chun cabhrú le hasbhaint téarmaíochta ó chorpais agus ó chuimhní aistriúcháin.

Acmhainní Comhréire agus Séimeantaice – Forbairtí Riachtanacha Amach Anseo

Tá an méid seo i measc cuid de na riachtanais is práinní:
- Graiméir ríomhaireachta a fhorbairt ag baint úsáid as creataí comhréire teoiriciúla ríomhaireachtúla seanbhunaithe amhail Graiméar Feidhmiúil Léacsach, Graiméar Struchtúr Frása Ceanntiomáinte, Graiméar Teaglamach Catagóireach, nó a leithéid, chun tacú leis an teicneolaíocht pharsála.
- Foclóir fiúsachta níos cuimsithí, ag tógáil ar Fhoclóir Briathra Gaeilge (FBG) nuair is féidir, agus ag úsáid sonraí corpais níos nua.
- Acmhainn de chineál VerbNet Gaeilge a fhorbairt i gcomhar le forbairt ar acmhainní léacsacha amhail FBG agus LSG (féach Acmhainní Séimeantaice).
- Acmhainn de chineál FrameNet a fhorbairt a chomhlánódh foclóir fiúsachta (féach thuas), chun patrúin úsáide briathra a aithint, le húsáid i bparsáil comhréire/séimeantaice riailbhunaithe agus in eilimintí riailbhunaithe an mheaisínaistriúcháin.
- Tacaíocht agus nuashonruithe leanúnacha do gach líonra séimeantach amhail Líonra Séimeantach na Gaeilge (LSG) nó Wordnet Ilteangach (Gaeilge), lena n-áirítear fíorú de láimh ar na mapálacha a ghintear go huathoibríoch i líonraí séimeantacha d'fhonn an cruinneas a fheabhsú chun iontaofacht i bhfeidhmchláir úsáideora deiridh amhail teasárais a chinntiú.
- Tuilleadh taighde ar ionduchtú uathoibríoch líonraí séimeantaice a chuirfeadh feabhas ar na naisc atá ann cheana, agus a dhéanfadh níos éasca é iontrálacha nua a chur leo.
- Tuilleadh taisclanna sonraí nasctha d'ábhar Gaeilge amhail DBpedia as Gaeilge.
- Tá gá le hointeolaíochtaí agus tacsanomaíochtaí a chruthú ar féidir leo tacú le forbairt na ngraf eolais.
- Ní dhearnadh aon taighde go dtí seo don Ghaeilge i réimse fáis na mbunachar eolais. Tá buntaighde ag teastáil chun leas a bhaint as na deiseanna a ghabhann leis na huirlisí seo.

References

1. Ó Dónaill, N., Foclóir Gaeilge-Béarla. 1977: Baile Átha Cliath.

2. Ó Dónaill, N. and P. Ua Maoileoin, An Foclóir Beag. 1991: Baile Átha Cliath.

3. de Bhaldraithe, T., Foclóir Béarla-Gaeilge. 1959: Baile Átha Cliath.

4. Maldonado, A. and D. Lewis, Self-tuning ongoing terminology extraction retrained on terminology validation decisions., in Proceedings of the 12th International conference on Terminology and Knowledge Engineering 2016. 2016: Copenhagen, Denmark. p. 91-100.

5. Kipper, K., et al., A Large-scale Extension of VerbNet with Novel Verb Classes, in Proceedings of the 12th EURALEX International Congress. 2006, European Association of Lexicography: Torino, Italy. p. 173--184.

6. Kipper, K., et al., Extensive Classifications of English Verbs, in Proceedings of the 12th EURALEX International Congress. 2006: Torino, Italy.

7. Fellbaum, C., WordNet: An Electronic Lexical Database. 1998, Cambridge, Massachusetts: MIT Press / A Bradford Book.

8. Baker, C.F., C.J. Fillmore, and J.B. Lowe, The Berkeley FrameNet Project, in 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1. 1998, Association for Computational Linguistics: Montreal. p. 86----90.

9. Petruck, M.R.L. and V. Kordoni, Robust Semantic Analysis of Multiword Expressions with FrameNet, in Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), Lisbon, Portugal. 2015: Lisbon, Portugal.

10. Johannsen, A., H. Martínez Alonso, and A. Søgaard, Any-language frame-semantic parsing, in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015, Association for Computational Linguistics: Lisbon, Portugal. p. 2062--2066.

11. Wigger, A., Caint Chonamara: Bailiúchán Hartmann. 2000.

12. Institiúid Teangeolaíochta Éireann, Corpus Náisiúnta na Gaeilge. 2000.

13. Uí Dhonnchadha, E., Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar. 2009, Dublin City University.

14. Lynn, T., Irish Dependency Treebanking and Parsing. 2016, Dublin City University and Macquarie University. p. 174.

15. Sulger, S., Irish Clefting -- The LFG Perspective. 2009, Universit\"at Konstanz.

16. Lauren Cassidy, Teresa Lynn, James Barry, and Jennifer Foster. 2022. TwittIrish: A Universal Dependencies Treebank of Tweets in Modern Irish. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 6869–6884, Dublin, Ireland. Association for Computational Linguistics.

17. Lynn, T. and J. Foster, Universal Dependencies for Irish, in Proceedings of the Second Celtic Language Technology Workshop. 2016: Paris, France. p. 79--92.

18. Scannell, K., Líonra Seimeantach na Gaeilge. 2007.

19. O'Regan, J., K. Scannell, and E. Uí Dhonnchadha, lemonGAWN: WordNet Gaeilge as Linked Data, in Proceedings of the 5th Workshop on Linked Data in Linguistics: Managing, Building and Using Linked Language Resources, LREC. 2016: Portoro, Slovenia. p. 36--41.

20. Arcan, M., J. McCrae, and P. Buitelaar, Polylingual Wordnet. https://arxiv.org/list/cs/recent, 2019.

21. Walsh, A., T. Lynn, and J. Foster, Ilfhocail: A Lexicon of Irish MWEs, in Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019). 2019. p. 162–168.

Caibidil 5
Uirlisí Próiseála Teanga Nádúrtha

5.1 Céard is Próiseáil Teanga Nádúrtha ann?

Próiseáil Teanga Nádúrtha (PTN) is ea an téarma a úsáidtear le cur síos ar an bpróiseáil uathoibríoch a dhéanann ríomhaire ar theanga an duine. Áirítear leis an bpróiseáil an iarracht chun ciall a bhaint as an teanga scríofa nó an teanga labhartha, agus áirítear léi freisin teanga an duine a ghiniúint mar fhreagra ar iarratas, mar chuid de dhialóg, nó mar aistriúchán. Is feiniméan ríchasta í teanga an duine; mar sin de, is gnách go ndéantar próiseáil uathoibríoch na teanga a bhriseadh síos ina céimeanna beaga iomadúla. Tugtar faoi gach ceann de na céimeanna sin trí bhíthin úsáid a bhaint as bogearraí atá dírithe ar thasc faoi leith. Uirlisí Próiseála Teanga Nádúrtha is ea a thugtar de ghnáth ar na bogearraí trína gcuirtear na tascanna sin i gcrích. Is iad na huirlisí sin, mar aon le hacmhainní teanga amhail corpas (Caibidil 3) agus bunachair eolais (Caibidil 4), na dúshraitheanna bunriachtanacha ar a leagtar na córais mhóra chasta a dtugtar cuntas mion orthu sa phlean seo. Léirítear roinnt de na hacmhainní agus de na huirlisí is tábhachtaí a úsáidtear sa Phróiseáil Teanga Nádúrtha i bhFíor 5.1.

5.2 Cén tábhacht atá le PTN agus cé dó é?

D'fhonn gur féidir gnáth-thascanna próiseála focal a dhéanamh amhail seiceáil litrithe agus gramadaí, ní mór a bheith in ann anailís a dhéanamh ar fhocail, ar nathanna agus ar abairtí an téacs, agus ní mór go mbeimid in ann moltaí cearta léacsacha agus gramadaí a thabhairt. Maidir le córais FRT le haghaidh an oideachais de, ní mór go mbeimid in ann measúnú a dhéanamh ar an ionchur teanga agus freagairtí cuí a ghiniúint. D'fhonn go mbeifear in ann modh so-úsáidte cumarsáide duine le ríomhaire a sholáthar, is amhlaidh a bheidh réimse leathan uirlisí PTN ag teastáil le haghaidh na hanailíse agus na giniúna moirfeolaíochta, clibeáil roinne cainte (RC), parsáil éadomhain, parsáil comhréire agus séimeantaice, srl., cúraimí a dtabharfar cuntas níos mine orthu sa chéad mhír eile. Ar cheann de na feidhmeanna tosaigh a baineadh as PTN le haghaidh na Nua-Ghaeilge, bhí corpas de théacsanna Gaeilge a anótáil le clibeanna roinne cainte ar mhaithe le forbairt an Fhoclóra Nua Béarla-Gaeilge.

<table>
<tr><td colspan="6">FEIDHMCHLÁIR: BOGEARRAÍ PROFÁLA CÁIPÉISÍ, FOGHLAIM RÍOMHCHUIDITHE TEANGA, MEAISÍNAISTRIÚCHÁN ACHOIMRIÚ, FREAGAIRT FIOSRUITHE, SIMPLIÚ TÉACS, SRL.</td></tr>
<tr><td colspan="5">PARSÁLAITHE</td><td>BAINC THAIRISCEANA</td></tr>
<tr><td colspan="4">PARSÁLAITHE ÉADOIMHNE</td><td>CORPAIS PHARSÁILTE</td><td>GRAIMÉIR RÍOMHAIREACHTÚLA</td></tr>
<tr><td colspan="3">CLIBEÁLAÍ ROINNE CAINTE</td><td colspan="2">RIALACHA STRUCHTÚIR FRÁSAÍ</td><td>AONÁIN AINMNITHE</td></tr>
<tr><td colspan="2">BOGEARRAÍ MOIRFEOLAÍOCHTA</td><td>MORFACHOMHRÉIR</td><td>MINICÍOCHTAÍ LÉACSACHA</td><td>STRUCHTÚR BRIATHRA</td><td>AICMÍ AINMFHOCAL</td></tr>
<tr><td>FOCLÓIRÍ DIGITEACHA</td><td>GRAIMÉAR TAGARTHA</td><td colspan="2">CORPAIS</td><td>LOGAINMNEACHA</td><td>AINMNEACHA PEARSANTA</td></tr>
<tr><td colspan="6">EOLAS TEANGA AGUS TEANGEOLAÍOCH</td></tr>
</table>

Fíor 5.1 Acmhainní agus Uirlisí PTN

5.3 Conas a oibríonn sé?

D'fhonn abairt a phróiseáil, d'fhéadfadh sé go n-áireofaí le gnáth-thascanna PTN an abairt a roinnt ina focail aonair (agus i roinnt aonaid ilfhoclacha amhail logainmneacha srl.), catagóirí roinne cainte a shannadh do na focail ionas go ndéanfar amach na hainmfhocail agus na briathair srl., agus na gnéithe moirfeolaíochta agus séimeantaice a shannadh do na focail ansin, .i. an uimhir uatha/iolra a shannadh do na hainmfhocail, an aimsir a shannadh do na briathair, srl. Ina dhiaidh sin, ní mór go rangófar na focail de réir nathanna agus go ndéanfar amach na gaolta idir nathanna ionas go dtiocfar ar an mbunchiall, .i. róil séimeantaice na n-aonán atá i gceist, rud a léiríonn 'cé a rinne gníomh agus cé air a ndearnadh é'. Tá de dhíth ar an ord imeachtaí sin, faoi mar a léirítear i bhFíor 5.2 é, roinnt uirlisí éagsúla PTN, a n-áirítear leis de ghnáth téacschomharthóir, anailíseoir moirfeolaíochta, parsálaí éadomhain, aitheantóir aonáin ainmnithe, lipéadóir róil shéimeantaigh agus parsálaí. Cuireann gach uirlis acu siúd faisnéis bhreise leis an abairt tosaigh ghnáth-théacs i riocht 'anótálacha' nó 'lipéad'.

Fíor 5.2 Anótáil PTN ar théacs le haghaidh feidhmchlár teanga

Modh próiseála atá bunaithe ar rialacha is ea an modh sin tríd is tríd. Nuair a bhíonn dóthain acmhainní ar fáil, is amhlaidh is féidir tairbhe mhór a bhaint as modhanna staitisticiúla agus meaisínfhoghlama chomh maith céanna. Mar shampla, nuair atáthar tar éis corpas sách mór de théacsanna anótáilte (nó corpais choibhéiseacha de ghnáth-théacsanna ailínithe i gcás an aistriúcháin de) a bhailiú agus tar éis dearbhú cáilíochta a dhéanamh orthu freisin, is amhlaidh is féidir modhanna agus uirlisí meaisínfhoghlama a úsáid, i riocht agus gur féidir leis an gcóras tosú ar phatrúin a aithint agus rialacha a bhaint as na patrúin siúd go huathoibríoch. Mar gheall ar dhul chun a rinneadh ar an domhainfhoghlaim le deireanas, is amhlaidh atáthar ag baint úsáid as leabuithe focail réamhthraenáilte (léirithe matamaiticiúla ar fhocail), rud a tugann léargas ar úsáid na teanga i gcomhthéacs agus a chuireann feabhas mór ar an bhfeidhmíocht dá réir sin.

5.4 A bhfuil déanta i leith na Gaeilge go dtí seo

Uirlisí Moirfeolaíochta

Áirítear le huirlisí moirfeolaíochta uirlisí le haghaidh téacschomharthú, leamú, anailís agus giniúint mhoirfeolaíochta. Is ionann téacschomharthú agus an tasc i dtaca le habairt nó le hurlabhra a roinnt in aonaid a dtugtar 'comharthaí' orthu, lena n-áirítear focail, poncaíocht agus míreanna ilfhoclacha, faoi mar is ionann leamú agus an tasc i dtaca leis an gceannfhocal foclóra a bhaineann le foirmeacha infhillte a aithint , .i. is é cuir an leama le haghaidh cuirfidh agus cuireann srl. Tá fáil ar anailíseoir agus gineadóir moirfeolaíochta staide críochta atá bunaithe ar rialacha le haghaidh na Gaeilge [1], ina n-úsáidtear rialacha moirfeolaíochta atá ionchódaithe de láimh agus atá bunaithe ar pharaidímí gramadaí [2, 3] agus ina bhfuil léacsacan mór inmheánach lena n-áirítear na ceannfhocail agus na leaganacha malartacha atá le fáil i bhfoclóir Uí Dhónaill [4]. Aithníonn na huirlisí moirfeolaíochta an leama, an chatagóir roinne cainte agus gnéithe bainteacha gach comhartha sa téacs. Is iomaí focal a bhféadfadh ní ba mhó ná aon roinn cainte amháin agus/nó aon leama amháin a bheith i gceist leo, agus déanann an t-anailíseoir moirfeolaíochta gach anailís fhéideartha i leith gach focail a aschur. Tá comhpháirt eile ag baint leis na huirlisí seo ina ndéantar an chatagóir roinne cainte is dóchúla le haghaidh focal anaithnid a thuar, bunaithe ar thréithe seachtracha na bhfocal féin, .i. réimíreanna nó iarmhíreanna inaitheanta nó foircinn leathana/chaola. Mar shampla, más focal dar críoch –faidh focal anaithnid, tuarfaidh an t-anailíseoir moirfeolaíochta gur briathar san aimsir fháistineach atá ann, fiú is nach bhfuil an fhréamh ná an leama le fáil sa léacsacan.

I gcás uirlisí dá leithéid atá bunaithe ar rialacha, tá sé tábhachtach go gcoinnítear an léacsacan cothrom le dáta. Bíonn focail agus téarmaíocht nua á dtabhairt isteach sa teanga go rialta (féach Caibidil 3 maidir le corpais agus Caibidil 4 maidir le foclóirí agus bunachair eolais foclóireachta) agus is gá dá réir sin go ndéantar nuashonrú rialta ar an léacsacan trí bhíthin focail agus téarmaíocht nua a chur isteach ann. Is féidir an léacsacan moirfeolaíochta a fheabhsú trí bhíthin tuilleadh míreanna ilfhoclacha a chur leis, .i. ar cosa in airde 'at a gallop', agus nathanna amhail ar muin na muice 'on the pig's back'.

ar	is	Copula
ar	ar	Noun+Masc+Com+Sg
ar	ar	Verb+PastInd
ar	ar	Prep+Simp
ar	ar	Particle+Verbal+Q+Past
ar	ar	Particle+Verbal+Rel
muin	muin	Noun+Fem+Com+Sg
na	na	Article+Pl+Def
na	na	Article+Gen+Sg+Def+Fem
muice	muc	Noun+Fem+Gen+Sg

Fíor 5.3 Anailís mhoirfeolaíochta ar "ar muin na muice" ina dtaispeántar an téacschomhartha i gcolún 1, an leama i gcolún 2, agus an chatagóir + mionghnéithe roinne cainte i gcolún 3

I staidéar Lynn et al.[5] ar úsáid na Gaeilge ar Twitter baineadh leas as leimeoir staitistiúil nó leimeoir atá bunaithe ar shonraí, Morfette [6]. Trí bhíthin an uirlis a thraenáil i leamaí (fréamhacha) d'inneachair tvuíteanna a deimhníodh de láimh, is amhlaidh a d'fhoghlaim an uirlis an chaoi chun na leamaí i mbreis agus 97% d'fhocail tvuíteanna nach bhfacthas roimhe sin a thuar. Tá gá le turgnaimh eile ina ndéantar anailís staitistiúil le huirlisí moirfeolaíochta atá bunaithe ar shonraí, amhail Morfette nó MARMOT [7], ar thacar mór sonraí de théacs caighdeánach Gaeilge agus de théacs torannach a ghin úsáideoir, .i. sonraí meán sóisialta.

Uirlisí Clibeála Roinne Cainte

Is é is cuspóir le clibeálaí roinne cainte (RC) ná an chlib chuí POS agus na gnéithe cuí eile a shannadh do gach focal in abairt. An deacracht a bhaineann le clibeáil RC is ea gur féidir le formhór na bhfocal feidhmiú ar bhealaí éagsúla, .i. d'fhéadfadh sé go mbeadh feidhm aidiachta le glan sa nath seomra glan nó d'fhéadfadh sé go mbeadh feidhm bhriathartha leis sa nath ghlan sé é. Cuirtear san áireamh i gclibeálaí RC atá bunaithe ar rialacha [8] na roghanna éagsúla anailíse a sholáthraíonn an t-anailíseoir moirfeolaíochta (féach thuas), agus cuirtear rialacha comhréire a códaíodh de láimh i bhfeidhm d'fhonn catagóir chuí RC an fhocail i gcomhthéacs faoi leith a dhéanamh amach, ar a mbíonn toradh dá leithéid a léirítear i bhFíor 5.4.

ar	ar	Prep+Simp
muin	muin	Noun+Fem+Com+Sg
na	na A	rticle+Gen+Sg+Def+Fem
muice	muc	Noun+Fem+Gen+Sg

Fíor 5.4 Leamú agus clibeáil RC le haghaidh "ar muin na muice"

Sa chás nach féidir leis an gcóras atá bunaithe ar rialacha teacht ar chinneadh atá bunaithe ar an gcomhthéacs logánta comhréire amháin, is amhlaidh is féidir úsáid a bhaint as faisnéis um minicíocht léacsach chun an chatagóir RC is dóchúla a dhéanamh amach. Mar shampla, in abairt faoi leith, ina bhféadfadh sé gur forainm a bhfuil ciall an tríú pearsa uatha leis é an focal siad nó gur ainmfhocal é a bhfuil an chiall 'meall nó 'at' leis, is amhlaidh a thabharfar tosaíocht do chiall níos minice an fhorainm siad, bunaithe ar an bhfaisnéis um minicíocht léacsach a baineadh as corpas Gaeilge (Caibidil 3). Cruthaíodh cheana féin go mbíonn rath ar chuir chuige hibrideacha i leith modhanna rialach agus staitistiúla ó thaobh uirlisí a fhorbairt le haghaidh teangacha eile [9]. Chomh maith leis sin, nuair a bheidh teacht ar dhóthain sonraí clibeáilte POS a deimhníodh de láimh, is amhlaidh a d'fhéadfaí clibeálaí roinne cainte don Ghaeilge a oiliúint ar lorg na sonraí clibeáilte POS atá ar shárchaighdeán (féach Caibidil 3).

Rinneadh roinnt turgnamh sa chlibeáil roinne cainte staitistiúil ar shonraí Twitter trí bhíthin anailíseoir staitistiúil a oileadh ar chorpas beag tvuíteanna [5]. Buntáiste a bhaineann leis an oiliúint sainréimse sin is ea gur fusa próiseáil a dhéanamh ar theanga neamhchaighdeánach, .i. sonraí Twitter, dá barr.

Parsálaithe Éadoimhne agus Uirlisí Atá Bunaithe ar Fhrása

Is éard is cuspóir le parsálaí éadomhain ná focail a ghrúpáil i bhfrásaí trí bhíthin an ceannfhocal agus na focail atá spleách ar an gceannfhocal a aithint, .i. i bhfrása ainmfhocail, beidh ceann-ainmfhocal agus d'fhéadfadh sé go mbeadh focail éagsúla eile ann atá spleách ar an ainmfhocal sin, amhail alt, nó aidiachtaí nó ainmfhocal cáilithe. Bíonn ról gramadaí ag frásaí nó smutáin, seachas ag focail aonair, san abairt, .i. is mar ábhar gramadaí nó 'cuspóir díreach', srl. a fheidhmeoidh an frása ainmfhoclach . Tá fáil ar pharsálaí éadomhain spleáchais atá bunaithe ar rialacha (páirtpharsálaí) le haghaidh na Gaeilge [8], ina ndéantar frásaí agus na feidhmeanna gramadaí atá leo (.i. ainmní, cuspóir díreach srl.) a aithint. Chun tógáil ar an méid sin agus parsálaí spleáchais comhréire atá bunaithe ar rialacha agus atá lánchuimsitheach a fhorbairt, beidh taighde breise le déanamh ar an struchtúr briathar-argóint [10] agus ar thréithe séimeantacha na n-ainmfhocal, (féach Caibidil 3 agus Caibidil 4).

Fíor 5.5 Abairt ag a bhfuil frásaí (nó smutáin) idir lúibíní maille le feidhmeanna agus spleáchais ghramadaí

```
Bhí na buaiteoirí ar muin na muice.

[S   [V Bhí bí+Verb+VI+PastInd+Len+@FMV]
     [NP na na+Art+Pl+Def+@>N
          buaiteoirí buaiteoir+Noun+Masc+Com+Pl+@SUBJ]
     [PP ar ar+Prep+Simp+@PP_SC
          [NP muin muin+Noun+Fem+Com+Sg+@P<
               na na+Art+Gen+Sg+Def+Fem+@>N
               muice muc+Noun+Fem+Gen+Sg+@N] ] ]
```

Parsálaithe Comhréire

Tugtar anailís chomhréire lánchuimsitheach ar abairt le parsálaí. Aithnítear inti gach gaol idir frásaí agus an ról a bhíonn acu maidir le cur síos a dhéanamh ar theagmhas. Forbraíodh parsálaithe comhréire staitistiúla lánchuimsitheacha le haghaidh na Gaeilge [11], ina nasctar le chéile gach focal de chuid na habairte de réir ról suntasach gramadaí a dtugtar gaolta spleáchais orthu, féach Fíor 5.6. Bíonn na parsálaithe comhréire seo bunaithe ar shonraí agus i dtuilleamaí sonraí as an Irish Dependency Treebank agus as an Irish Universal Dependency Treebank (féach Caibidil 3.4) chun struchtúr gramadaí na teanga a spreagadh agus a fhoghlaim. Fíor 5.6 Abairt ag a bhfuil crann spleáchais comhréire ina dtaispeántar feidhmeanna agus spleáchais ghramadaí

Mar sin de, is féidir go dtuarfaidh na parsálaithe, bunaithe ar an dóchúlachta a bhaineann le meascán tréithe sa chrannbhanc (foirmeacha focail, leamaí, gnéithe moirfeolaíochta, cineálacha spleáchais a chonacthas cheana) struchtúr comhréire abairtí nach bhfacthas roimhe sin. Mar is gnách i gcás uirlisí atá bunaithe ar shonraí de, is amhlaidh a ghabhfaidh cruinneas an pharsálaithe i méid faoi mar a thiocfaidh fás ar na crannbhainc .

Figure 5.6 Sentence with syntactic dependency tree showing grammatical functions and dependencies

D'fhéadfadh sé go mbeidh tairbhe le baint as taighde ar mhodhanna eile parsála seachas an pharsáil spleáchais (.i. parsáil comhpháirte, parsáil de cuid na Gramadaí Catagóirí Teaglamaí (Combinatory Categorial Grammar, GCT feasta)) maidir le fothascanna a bhítear ag iarraidh a réiteach d'fhonn tascanna níos príomha a réiteach. Chuige sin, d'fhéadfadh sé go mbeadh réamhobair tosaigh GCT le haghaidh Ghaeilge na hAlban [12] ina cuidiú le hiarrachtaí taighde i dtaca le crannbhanc na Gaeilge de.

5.5 Conairí Taighde

Chomh maith leis an bhforbairt agus leis an gcothabháil leanúnach atáthar a dhéanamh ar uirlisí ar tugadh cur síos orthu i míreanna roimhe seo, déantar forbhreathnú sna míreanna seo a leanas ar roinnt de na huirlisí PTN atá ar fáil le haghaidh teangacha eile, ach uirlis nár forbraíodh le haghaidh na Gaeilge go fóill. Tugaimid achoimre ar a dtábhacht agus ar a n-ábharthacht maidir lena bhfeidhmeanna don úsáideoir deiridh. Ós réimse é seo atá ag athrú go mear, féadaimid a bheith ag súil leis go dtiocfaidh uirlisí nua agus modhanna nua oibre chun cinn sna blianta amach romhainn. Tá sé bunriachtanach dá réir sin go bhfanfaidh lucht na Gaeilge cothrom le dáta i dtaca leis na forbairtí atá á ndéanamh go hidirnáisiúnta.

Lipéadú Róil Shéimeantaigh agus Parsáil Shéimeantach

Déanann parsálaí struchtúr comhréire agus róil ghramadaí a shannadh don abairt, ach, i gcás tascanna go leor de, is gá go mbeifear ar an eolas faoi na róil shéimeantacha a bhaineann leis na haonáin san abairt d'fhonn go mbeifear ar an eolas faoi "cé a rinne gníomh agus cé air a ndearnadh an gníomh". I gcás go leor feidhmchlár tuisceana teanga nádúrtha amhail asbhaint faisnéise agus achoimriú , ní mór go mbeidh rochtain acu siúd ar fhreagairt ceisteanna agus anailís dearcaidh srl.., sin ní mór go mbeidh rochtain acu ar a leithéid d'fhaisnéis séimeantach.

(1) Thug Seán leabhar dá chara
Gave Seán book to-his friend
AGENT THEME RECIPIENT
REL ARG0 ARG1 ARG2
'Seán gave a book to his friend'

Is iondúil go ndéanann lipéadóir róil shéimeantaigh (LRS) róil shéimeantacha a shannadh do na hargóintí agus do na maolaitheoirí briathair. Mar shampla, i sampla (1) 'gníomhaí' is ea an ról atá ag Seán, i.e. 'an té a dhéanann an gníomh', 'faighteoir' is ea an de ról atá ag chara agus 'téama' is ea an ról atá ag leabhar. Nuair a bhíonn ríomhaire in ann na comhpháirteanna sin san abairt a dhéanamh amach [13, 14], is amhlaidh is fearrde é chun tascanna amhail an t-aistriúchán, an t-achoimriú, brath earráidí agus mar sin de, a dhéanamh.

D'fhéadfadh sé go mbeadh an obair a rinneadh ar fhiúsacht an bhriathair i bhFoclóir Briathra Gaeilge (FBG) (féach Caibidil 4) ina bonn le huirlis lipéadúcháin róil shéimeantaigh a fhorbairt.

Uathimdhealú chiall an fhocail

Is tasc bunriachtanach é uathimdhealú chiall an fhocail (UCF) i bhfeidhmchláir go leor a bhaineann leis an asbhaint faisnéise, an meaisínaistriúchán, srl. Bíonn níos mó ná aon chiall amháin le focail go leor agus is é an dúshlán a bhíonn le sárú ná an chiall cheart a roghnú i gcomhthéacs faoi leith. Mar shampla, tá roinnt cialla éagsúla leis an bhfocal caith, .i. d'fhéadfadh sé go mbeadh an chiall `throw', `spend', `wear', nó `smoke' an Bhéarla leis. I bhfeidhmchláir PTN, amhail an asbhaint faisnéise, an t-achoimriú agus innéacsú doiciméad, agus an meaisínaistriúchán srl., tá sé bunriachtanach go n-aithneofar an chiall chuí ionas gur féidir freagraí cruinne a thabhairt ar fhiosrú.

Tá méadú suntasach tar éis teacht ar an úsáid a bhaintear as WordNets [15] agus ar an úsáid a bhaintear as líonraí néaracha agus as leabuithe focal [16] i gcás uirlisí PTN mar a mbaintear leas as modhanna meaisínfhoghlama. Ón taobh sin de, d'fhéadfadh sé go mbeadh na hacmhainní séimeantacha, lena n-áirítear Líonra Séimeantach na Gaeilge (LSG) (féach Caibidil 4) mar dhúshraith arbh fhéidir uirlisí UCF úrscothacha, cosúil leo siúd atá in úsáid i dteangacha eile faoi láthair, a fhorbairt le haghaidh na Gaeilge.

Aithint aonáin ainmnithe

Is minic gur ainmfhocal dílis trína ndéantar tagairt do dhuine, d'áit nó d'eagraíocht iad na hainmfhocail anaithnide i dtéacs. In aitheantóir aonáin ainmnithe (AAA) aithnítear agus rangaítear focail agus frásaí ina n-ainmneacha ar eagraíochtaí, ar dhaoine, ar áiteanna, ar ghnéithe ama agus mar sin de.

(2) Bunaíodh [Grúpa Gnímh an N59] san oíche [Dé Céadaoin]
ORG DATE
established [group action-GEN the N59] in-the night [Wednesday]
'The N59 Action Group was established Wednesday night'

Bíonn uirlis AAA an-tairbeach le haghaidh tascanna amhail an t-achoimriú téacs, an aisghabháil faisnéise agus an t-aistriúchán [17, 18]. Go dtí seo, ní raibh fáil ar aon chóras aitheanta aonáin ainmnithe le haghaidh na Gaeilge. Tá roinnt acmhainní bunúsacha (liostaí d'aonáin ainmnithe) ar fáil trí theicneolaíocht clibeála roinne cainte, agus trí logainmneacha ar logainm.ie, ach tá go leor oibre ag teastáil go fóill ionas go mbeidh uirlis chuimsitheach AAA le fáil ón taighde sin.

Anailís seintiminte

Is éard is ciall le seintimint ná an chiall nó an mothúchán atá á c(h)ur in iúl i dtéacs nó in urlabhra. Cuir i gcás, seintimintí amhail déistin, fearg, ionadh nó lúcháir. In uirlisí anailíse seintiminte, is minic a bhítear ag díriú ar an tseintimint atá á cur in iúl, bíodh an tseintimint sin dearfach, diúltach nó neodrach. Is féidir feidhm a bhaint as teicneolaíocht dá leithéid i gcomhthéacsanna éagsúla, mar shampla tomhas a dhéanamh ar a mhéad agus a thacaíonn an pobal leis an rialtas maidir le hábhar faoi leith, nó measúnú a dhéanamh ar thuairim an mhargaidh de tháirgí nó de sheirbhísí gnólachta. Sa lá atá inniu ann, is minic gur chun na críche sin [19] a dhéantar an anailís seintiminte ar inneachar ar líne.

Mar sin féin, is féidir go mbeidh an anailís seintiminte ina cuidiú le taighdeoirí i réimse na ndaonnachtaí digiteacha, mar a ndéanann taighdeoirí staire cuardach ar ailt nó ar cháipéisí ina ndéantar tuairisc nó léiriú dearfach nó diúltach ar ábhar.

De thoradh tionscadal anailíse seintiminte a rinne taighdeoirí ó DCU, tugadh léiriú ar dhearcadh an phobail sa tréimhse díreach roimh olltoghchán 2016; is trí anailís ar tvuíteanna a rinneadh an taighde sin [20].

Simpliú téacs

Is éard is cuspóir le huirlis simpliúcháin téacs ná anailís a dhéanamh ar théacs i riocht agus go ndéanfar an téacs céanna níos simplí agus níos soléite don léitheoir, nuair is féidir. I gcás teicnící simpliúcháin téacs de, dírítear ar an léacsacan agus/nó an struchtúr comhréire a shimpliú. Is iomaí feidhm a fhéadtar a bhaint as an simpliú téacs i réimse an oideachais [21, 22, 23] agus an riaracháin phoiblí [24]. Baineann uirlisí dá leithéid go mór le hábhar i gcomhthéacs téacsanna a chur ar fáil d'fhoghlaimeoirí Gaeilge sa chóras oideachais, agus i gcomhthéacs cáipéisí a sholáthar do dhaoine sa phobal agus iad ag plé leis an rialtas áitiúil agus lárnach.

5.6 Moltaí

- Tá cothabháil agus nuashonrú rialta ar uirlisí atá ann cheana féin de dhíth d'fhonn coinneáil cothrom le dáta leis an teicneolaíocht úrscothach a bhíonn ag athrú de shíor, lena n-áirítear uasdátú bogearraí.
- Ba chóir don fhorbairt ar uirlisí nua a bheith á stiúradh ag taighde agus ag an bhfeidhm a bheidh le baint astu chun go gcinnteofar go mbeidh siad chomh hábhartha agus chomh hin-úsáidte agus is féidir.
- Ba chóir go mbeadh maoiniú ar fáil don óstáil agus don chothabháil leanúnach ar comhéadain úsáideoirí chun go gcinnteofar an teacht is mó agus is féidir ar na hacmhainní luachmhara seo.
- Is gá taisce lárnach de shonraí léacsacha agus téarmeolaíochta atá i bhformáid choiteann a bhunú (féach Caibidil 4), ionas gur féidir leis forbróirí na huirlisí PTN a choinneáil cothrom le dáta ar bhealach tráthrialta solúbtha.
- Cuid mhaith de na huirlisí PTN a ndéantar cur síos orthu sa chaibidil seo, idir sheanbhunaithe agus nua, tá siad i dtuilleamaí na mbonn teoirice a luaitear i gCaibidil 2 agus fáil a bheith ar bhunachair eolais léacsaigh agus ar chorpais atá mór agus cruinn a ndóthain, mar a mhínítear i gCaibidil 3 agus i gCaibidil 4.
- Tá gá le corpais níos mó ar shárchaighdeán ionas gur féidir feabhas a chur ar an gclibeáil roinne cainte agus ar an bparsáil a dhéantar ar inneachar foilsithe agus ar inneachar arna ghiniúint ag úsáideoirí amhail inneachar Facebook, Twitter srl. (féach Caibidil 3).
- Tá feabhsú na parsála spleáchais atá bunaithe ar rialacha le haghaidh Gaeilge i dtuilleamaí fhorbairt acmhainní léacsacha, amhail foclóir fiúise briathair agus faisnéis maidir le rangú séimeantaice na an ainmfhocail (féach Caibidil 4).
- Tá feabhsú na parsála spleáchais staitistiúla le haghaidh na Gaeilge i dtuilleamaí crannbhainc na Gaeilge a leathnú (féach Caibidil 3) agus i dtuilleamaí teicnící feabhsaithe parsála chomh maith.
- Tá acmhainní léacsacha agus corpais anótáilte ag teastáil (féach Caibidil 3 agus Caibidil 4) chun uirlisí nua PTN a fhorbairt le haghaidh lipéadú róil shéimeantaigh (LRS) agus uathimdhealú chiall an fhocail (UCF) .
- Tá parsálaithe éadoimhne/parsálaithe comhréire chomh maith le hacmhainní foclóireachta lena n-áirítear logainm.ie le haghaidh logainmneacha Gaeilge ag teastáil chun uirlisí aitheanta aonáin ainmnithe (AAA) a fhorbairt.
- Ionas gur féidir uirlisí seintiminte a fhorbairt, is gá leathadh a dhéanamh ar léacsacan reatha na bhfocal a bhfuil lipéad polaraíochta orthu [20] agus is gá an léacsacan a fhairsingiú tuilleadh ionas go dtiocfaidh réimsí eile faoina scóip.
- D'fhonn uirlisí simpliúcháin téacs a fhorbairt, is gá feidhm a bhaint as teoiric na comhréire (Caibidil 2) mar aon le hacmhainní léacsacha agus comhréire (Caibidil 4 agus an chaibidil seo).

References

1. Uí Dhonnchadha, E., An Analyser and Generator for Irish Inflectional Morphology using Finite-State-Transducers, in School of Computing. 2002, Dublin City University: Dublin, Ireland.

2. An Gúm, Graiméar Gaeilge na mBráithre Críostaí. 2 ed. 1999, Baile Átha Cliath: An Gúm.

3. Christian Brothers, New Irish Grammar. 1988, Dublin: C J Fallon.

4. Ó Dónaill, N., Foclóir Gaeilge Béarla. 1977, Baile Átha Cliath: Oifig an tSoláthair.

5. Lynn, T., Kevin Scannell and Eimear Maguire (2015), Minority Language Twitter: Part-of-Speech Tagging and Analysis of Irish Tweets, In Proceedings of the First Workshop on Noisy User-generated Text, Beijing, China

6 Chrupała, G., Dinu, G., van Genabith J. 2008, Learning Morphology with Morfette. Proceedings of LREC 2008.

7. Müller T, Schmid H, and Schütze H . 2013, Efficient higher-order crfs for morphological tagging. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)

8. Uí Dhonnchadha, E., Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar, in School of Computing. 2009, Dublin City University: Dublin, Ireland.

9. Loftsson, H., Helgadóttir, S., and Rögnvaldsson, E. (2011). Using a morphological database to in-crease the accuracy in pos tagging. In: Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP), pages 49–55, Hissar, Bulgaria.

10. Wigger, A. (2008). Advances in the lexicography of Modern Irish verbs. In: Bloch-Rozmej, A. (ed.): Issues in Celtic Linguistics pp. 233-250.

11. Lynn, T., Irish Dependency Treebanking and Parsing. 2016, Dublin City University and Macquarie University.

12. Batchelor, C. 2014. gdbank: The beginnings of a corpus of dependency structures and type-logical grammar in {S} cottish Gaelic, in Proceedings of CLTW2014, First Celtic Language Technology Workshop XVII, Dublin, August.

13. Gildea, D., & Jurafsky, D. (2002). Automatic Labeling of Semantic Roles. Computational Linguistics 28(3), 245-288.

Caibidil 6
Giniúint Teanga Nádúrtha

6.1 Céard is Giniúint Teanga Nádúrtha ann?

Is éard is Giniúint Teanga Nádúrtha (GTN) ann ná an réimse den Phróiseáil Teanga Nádúrtha a bhaineann le teanga nádúrtha a ghiniúint as foinsí sonraí nach mbaineann leis an teanga, amhail sonraí uimhriúla, ábhair leata shéimeantaice, amhail foirm loighce nó bunachair eolais struchtúrtha [1, 2]. Cé go bhféadfadh córas téacs-le-téacs a bheith i gceist le GTN .i., achoimriú coimritheach ar théacs nó tuairiscí téacsúla a ghiniúint as íomhá nó físeán [2], is ar an gcineál GTN ar a dtugtar córais Sonraí go Téacs [1] a dhírítear aird sa phíosa anseo. D'fhocail ghearra, is féidir linn córas GTN a thuiscint amhail is gur aistritheoir é a dhéanann sonraí (.i. sonraí uimhriúla as tuairisc aimsire nó as punann airgeadais) a thiontú go hachoimre teanga nádúrtha shaincheaptha (.i. tuairisc aimsire nó tuairisc shainoiriúnaithe airgeadais).

Sampla simplí is ea córas ina gcruthaítear litreacha. Is iondúil nach mbíonn rialacha gramadaí i gceist leo siúd, ach d'fhéadfadh sé go nginfidís litir chuig tomhaltóir, .i. ina ndeirtear gur sáraíodh an teorainn caithimh a ghabhann le cárta creidmheasa. Is é sin le rá, baintear úsáid as teimpléad i gcórais shimplí (nó éadoimhne) GTN, rud atá cosúil le postchumasc doiciméid Word, ach déantar an téacs a chruthú ar bhonn dinimiciúil i gcórais GTN níos casta agus níos nua. Ar aon dul le réimsí eile den teicneolaíocht teanga, is féidir an méid sin a dhéanamh trí úsáid a bhaint as múnlaí follasacha teanga (.i. graiméir nó tacair rialacha) nó as múnlaí staitistiúla a eascraíonn ó anailís ar shonraí reatha, .i. téacs a scríobh duine.

Agus parsáil ar siúl, ní mór don chóras imdhealú a dhéanamh ar an abairt ionchuir ionas go gcruthófar léiriúchán ar an teanga nádúrtha is strúchtúrtha ó thaobh meaisín de nó is soléite ag an duine. I gcás NLG de, ní mór don chóras cinntí a dhéanamh faoin gcaoi a gcuirfear focail ar choincheap.
Tá NLG fadbhunaithe ach níorbh é ach go dtí le gairid go raibh teacht go forleathan ar theicneolaíocht tráchtála NLG.

Giniúint Teanga Nádúrtha (GTN) i gcomparáid le Tuiscint Teanga Nádúrtha (TTN)

Is cosúil é an cúlra stairiúil a bhaineann le GTN agus macasamhail TTN agus an Mheaisínaistriúcháin (Caibidil 11), dá mhoille é, i ngeall ar fhorbairt iarrachtaí measúnaithe i measc lucht taighde, forbairt modhanna atá bunaithe ar shonraí agus soláthar chórais tráchtála GTN ar deireadh. Ba de thairbhe an bundúshlán a bhaineann le fadhb GTN í an mhoill úd. I gcás TTN de, is fearr mar a shainmhínítear an beart atá le déanamh agus tá eolas agus fáil fhorleathan ar an ionchur – an teanga nádúrtha – ar an nGréasán. Má tá féin, i dtaca le GTN de, tá eolas ar an t-aschur (an téacs), cé go raibh easnamh r ionchuir chuí (sonraí) atá feiliúnach le haghaidh giniúna [3] san aimsir a caitheadh. Dá réir sin, agus de bharr an easnaimh stairiúil ar chorpais fheiliúnacha sonraí-go téacs freisin, ba mhór an dúshlán é teacht ar chomhaontú faoi mhodhanna measúnaithe agus faoi chreatlaí GNT. Mar gheall ar fhorás an Olltiomsaithe Sonraí agus ar an mborradh atá tar éis teacht ar chuir chuige líonraí néaracha doimhne i leith GNT le deireanas, is amhlaidh atáthar ag cur spéis sa ghiniúint teanga an athuair, agus tá ag teacht chun cinn tacair sonraí sonraí-go-téacs [4, 5], ina mbaintear úsáid as leaganacha éagsúla de Léiriú Céille Teibí [LCT] mar fhormáid ionchuir aontaithe. Dá ainneoin sin, is iondúil go mbíonn a bhformhór de bhunghnéithe foinse oscailte NLG a bhfuil fáil orthu scaipthe, as feidhm nó faoi úinéireacht dhílsithe. Mar sin de, deis taighde a bhfuil gealladh fúithi is ea ardán comhchoiteann oscailte nó foireann uirlisí in-athúsáide le haghaidh GTN a fhorbairt, rud arbh fhéidir a bheith ina chuidiú leis an mbearna taighde idir TTN agus GTN a laghdú.

6.2 Cad chuige a bhfuil GTN tábhachtach agus cé dó a bhfuil sé tábhachtach?

I ré seo an Olltiomsaithe Sonraí, tá borradh agus biseach faoin réimse domhanda sonraí i gcónaí, agus tuartar go méadófar faoi dheich go 163 zeitibheart (zb, sin le rá, aon trilliún gigibheart) é faoin mbliain 2025 . Tá scrogall maidir le rochtain ar eolas, nach bhfuil aon dul as, le teacht ar an réimse is Anailísíocht Olltiomsaithe. Is cás sin a fhágann go mbainfear as an réimse léargas a bheidh brabúsach ach a bheidh dothuigthe don té nach saineolaí é chomh maith céanna. Tá sé áirithe go bhfuil an Ghiniúint Teanga Nádúrtha (GTN) ar cheann de na mílte deiridh atá le rith i rás an Olltiomsaithe Sonraí .

Ó thaobh cúrsaí tráchtála de, is iad córais sonraí-go-téacs a ghineann achoimrí téacsúla ar bhunachair agus ar thacair sonraí atá i bhfeidhmchláir GTN is rathúla go dtí seo; is iondúil go ndéantar anailís agus giniúint téacs sna córais. Go háirithe, forbraíodh an iomad córas ina gcruthaítear réamhaisnéis théacsúil aimsire as sonraí loma stáisiúin aimsire. Ar an gcóras ba thúisce dá leithéid a cuireadh i bhfeidhm bhí FoG [7], córas a bhí in úsáid ag Environment Canada chun réamhaisnéise théacsúil aimsire a ghiniúint i bhFraincis agus i mBéarla i mblianta tosaigh na 1990í. Spreagadh saothar eile, idir shaothar taighde agus saothar tráchtála, de bharr an rath a bhí ar FoG. Áirítear leis an taighde is déanaí sa réimse sin tionscnamh inar taispeánadh gur fearr le

húsáideoirí ar uairibh réamhaisnéis ríomhghinte aimsire ná an réamhaisnéis a scríobhann daoine, agus is cuid den chúis a bhí leis sin gur baineadh úsáid as téarmaíocht ní ba sheasmhaí i réamhaisnéis na ríomhairí [8,9] agus gur léiríodh gur féidir úsáid a bhaint as teicnící staitistiúla chun réamhaisnéis aimsire ar ardchaighdeán a ghiniúint. Áirítear leis na húsáideanna is déanaí a baineadh as FoG réamhaisnéis téacs-bhreisithe de chuid Met Office na Ríochta Aontaithe [10].

Tá an réimse is achoimriú téacs ann go mór i dtuilleamaí GTN. Feictear úsáid dá leithéid á baint i gcomhthéacs amhail achoimriú doiciméid (achoimre a dhéanamh ar alt nó ar thuairisc nuachta) chomh maith leis na blúiríní a chruthaítear i bhfeidhmchláir Aisghabhála Faisnéise (féach Caibidil 12), mar a soláthraítear d'úsáideoir torthaí ar chuardaigh chomh maith le hachoimre ghairid ar an doiciméid a aimsíodh (.i. leathanach gréasáin) ionas gur féidir leis an úsáideoir a shocrú an fiú cliceáil ar an nasc a chuirtear ar fáil nó nach fiú.

I gcomhthéacs na Gaeilge, is féidir le teicneolaíocht GTN bearna a laghdú i dtaca le hinneachar Gaeilge a chruthú ar ala na huaire lena úsáid trí mheáin dhigiteacha. Chonaiceamar thuas an úsáid rí-éifeachtach is féidir a bhaint as an teicneolaíocht sin chun réamhaisnéis aimsire a dhéanamh as sonraí loma. Is féidir le Met Éireann agus le soláthraithe seirbhíse eile a mbaineann a gcuid seirbhísí le sonraí, .i. tuairiscí tráchta, modhanna oibre den chineál céanna a chur in úsáid is féidir leo siúd leas a bhaint as na teicnící céanna ina gcuid oibre. Baineadh úsáid rí-éifeachtach as teicnící GTN chun tuarascálacha ar fhoinsí airgeadais agus ar fhoinsí sonraí struchtúrtha eile a chruthú go huathoibríoch. Ar lorg an mhúnla sin, is iomaí deis atá ann inneachar Gaeilge inar féidir freagairt fhíor-ama a thabhairt ar athrú tobann ar an mbunfhaisnéis a chruthú go mear. Ar an mbonn sin, is féidir fógraí seirbhíse poiblí, tuairiscí tuarascálacha agus foláireamh i gcásanna fáisnéis éigeandála a chur ar fáil láithreach bonn don dá phobal teanga in Éirinn gan gá leis an aistriú ná leis an logánú.

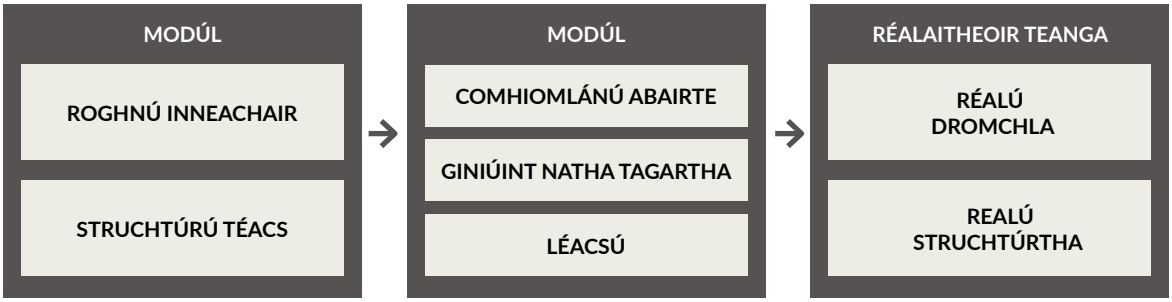
6.3 Conas a oibríonn sé?

D'fhéadfadh sé go mbeadh an próiseas na giniúna téacs, sa riocht is simplí de, baileach chomh simplí le hathúsáid a bhaint as athróga seanchaite téacs/stoc-athróga téacs a nasctar lena chéile le "gliú-théacs/téacs nasctha" statach d'fhonn inneachar nua a chruthú i bpróiseas atá cosúil leis an bpostchumasc (.i. [AINM] a chara, Fuaireamar d'iarratas ar chlárúcháin ar an [LL/MM/BBBB]). Is féidir go mbeidh toradh sásúil ar an gcur chuige sin i gcás úsáidí áirithe ach is i bhfad óna bheith sofaisticiúil, inscálaithe nó inaistrithe idir réimsí agus tascanna é. De ghnáth, bíonn céimeanna éagsúla den dearadh cliste ar a leithéid de chumasc faisnéise, i gceist le córas lándéanta GTN ionas gur féidir téacs atá líofa agus a bhfuil cuma nádúrtha air a ghiniúint.

Tascanna agus leagan amach GTN

Is féidir córas GTN sonraí go téacs a bhriseadh síos de réir tascanna éagsúla, a bhíonn le fáil go minic ina bhformhór de chórais GTN. Orthu sin tá:

- Sonrú Inneachair: Socrú ar an bhfaisnéis as an mbunachar ar cheart a áireamh sa téacs atá á ghiniúint; .i. gach a bhfuil suntasach. Cuir i gcás, i dtuairisc aimsire, d'fhéadfadh sé go mbeadh sonraí faoi bháisteach throm ina gcomhartha gur chóir rabhadh tuilte a ghiniúint.
- Struchtúrú Téacs (a dtugtar Pleanáil Doiciméid nó Téacs air freisin): Socrú ar an ord ina gcuirfear an fhaisnéis (ar a dtugtar teachtaireachtaí/teagmhais) i láthair sa téacs; Grúpáil na i dtéacs a gineadh (chur i gcás, iad a chur ina bparagraif).
- Comhiomlánú Abairte: Baineann an méid sin le socrú ar an bhfaisnéis ar chóir í a chur i láthair in aon abairt faoi leith; .i., socrú ar na teachtaireachtaí/teagmhais ar chóir iad a chumasc in aon abairt amháin. Is éard is cuspóir leis faisnéis chosúil a chumasc ionas go gcuirfear cosc leis an athrá agus go gcuirfear feabhas ar inléiteacht agus líofacht an téacs. Baineann an méid sin le comhábhair chomhréire (féach Caibidil 5) a chumasc, amhail abairtí agus frásaí.
- Léacsúchán: Na focail agus na nathanna cuí a roghnú chun faisnéis a chur in iúl; Baineann leis an méid sin na focail inneachair (ainmfhocail, briathra, aidiachtaí, agus dobhriathra) i dtéacs ginte a roghnú. Is éard a bhaineann leis an roghnú focail is simplí ná coincheap fearainn (a léirítear i bhfoirm ointeolaíochta seans) a mhapáil



Fíor1: Gnáthleagan amach thrí chéim GNT

ar fhocal. Cás níos casta ná sin is ea an tráth a gcuirtear coincheap faoi leith r in iúl trí úsáid a bhaint as focail nó parafrásaí éagsúla i gcásanna éagsúla faoi leith (.i. tairseach mhatamaiticiúil i gcomparáid le tairseach dorais).
- Giniúint natha tagartha: Na focail agus na frásaí a roghnú chun tagairt do nithe nó d'aonáin (.i. Na torthaí seo, an ráiteas sin, iad, é, í, srl.).
- Réalú Teanga: Tiontaítear léiriúcháin theibí (focail agus frásaí) go habairtí dea-chumtha de réir rialacha na comhréire, na deilbhíochta agus na hortagrafaíochta (Caibidil 2). Sa chéim seo, cuirtear lena chéile an Réalú Dromchla agus an úsáid chuí as marcáil (Réalú Struchtúrtha).

Tá i bhFíor 1 thuas forléargas ar an ngnáthleagan amach trí chéim [1], mar a n-eagraítear na tascanna GNT thuasluaite de réir trí mhodúl: Modúl Pleanála Téacs, Modúl Micreaphleanála agus Réalaitheoir Teanga. Tá cineálacha leagain amach eile ann, mar atá an cur chuige comhtháite, idirghníomhach, athbhreithnitheach nó cláir dhuibh [3]. D'fhéadfadh gur ar lorg stíl pleanála IS/AI a ghlactar cuir chuige eile i leith GNT nó cur chuige atá dírithe ar shonraí .i. cur chuige staitistiúil, stocastach nó meaisínfhoghlama [3]. I gcás roinnt córas de, d'fhéadfadh sé freisin go roinnfí tascanna thar modúil agus go mbeadh modh a n-eagraithe éagsúil. Is gnách gur córais chomhtháite ceann go ceann a bhíonn sna cuir chuige is déanaí atá bunaithe ar shonraí, go háirithe cinn domhainfhoghlama [4].

6.4 A bhfuil déanta i leith na Gaeilge go dtí seo

Go dtí seo, ní hann d'aon chóras neamhspleách aitheanta NLG amháin ná aon uirlis feidhme amháin le haghaidh na Gaeilge. Mar sin féin, tá cásanna úsáide agus úsáidí féideartha (faoi mar a pléadh thuas) ann agus is amhlaidh a thiocfaidh a leithéid chun cinn de réir mar a dhéantar dul chun cinn ar úsáid as an na teicneolaíochta inár saol laethúil. Níl acht ná amhras ann go bhféadfar tarraingt ar shaothar rathúil agus ar theicnící cruthaithe de chuid réimsí eile an ghoirt chun leas réimse seo na Teicneolaíochta Teanga a dhéanamh.

Tá baint nádúrtha ag obair an Mheaisínaistriúcháin (Caibidil 11) agus na Parsála (Caibidil 5) le réimse GNT. Agus sonraí téacs á bparsáil agus á dtiontú go léiriúcháin eile ag an bparsálaí, is amhlaidh a deirtear go bhfuil "eolas" á bhaint ag an bparsálaí céanna. Is féidir próiseas na parsála a chur in oiriúint d'aschur teanga nádúrtha a chruthú trí bhíthin tacar rialacha gramadaí le haghaidh foinse ionchuir faoi leith nó foinse sonraí eolais faoi leith a chur i bhfeidhm. Ar an gcuma chéanna, agus aistriúchán ó bhunteanga go spriocteanga á dhéanamh , ní mór go nginfidh an t-ineall MA, uair éigin, aschur sa spriocteanga atá ceart de réir an "eolais" a bhain sé as an bhfoinse ionchuir. Ar an gcaoi sin, d'fhéadfadh sé go mbeadh saothar MA ar an díchódú (ar lorg an mhodha atá bunaithe ar shonraí) agus saothar ar an iar-ghiniúint (ar lorg an mhodha atá bunaithe ar rialacha) ina mbunús leis an saothar tosaigh ar GTN Gaeilge. Agus de réir mar a dhéantar feabhsúcháin nua ar na réimsí cosúla sin de thairbhe Taighde agus Forbartha, is amhlaidh a d'fhéadfadh na feabhsúcháin cheánna dul chun tairbhe do chórais GTN.

6.5 Moltaí

- Moladh criticiúil a mbeadh tionchar ríthábhachtach aige ar an réimse is GTN na Gaeilge ann ba ea an réalaitheoir móréilimh atá bunaithe ar rialacha – is é sin SimpleNLG[11] – a chur in oiriúint don Ghaeilge. Tá an tóir ar an uirlis sin ag dul i méid i réimse an tionscail agus an léinn mar chur chuige tosaigh i leith tascanna giniúna sa chás go bhfuil easnamh acmhainní data2text le haghaidh teanga áirithe agus is féidir í a chur i bhfeidhm chun críche na Gearmáinise, na Spáinnise agus na hIodáilise. Is amhlaidh an scéal i gcás NLG de, agus is ar an údar sin gur ar an mBéarla amháin a cuireadh formhór na gcur chuige (domhainfhoghlama) nua atá bunaithe ar shonraí i bhfeidhm go dtí seo. Is féidir úsáid a bhaint freisin as aschuir atá bunaithe ar rialacha chun modhanna atá bunaithe ar shonraí a bhrógstrapáil.
- Dá ndéanfaí logánú Gaeilge ar SimpleNLG b'amhlaidh a chuirfí dlús le taighde bunúsach GTN chun craoltóireacht, tuairisciú nó fógraíocht seirbhíse poiblí uathoibrithe thar ceann an Stáit a chur i bhfeidhm, sa chaoi is nach ar phobal labhartha an Bhéarla in Éirinn amháin a dhéanfaí freastal a thuilleadh leis na seirbhísí seo.
- Tá dlúthbhaint ag an méid sin leis na moltaí Aisghabhála Faisnéise a dhéantar i gCaibidil 12 i ndáil le blúiríní d'achoimrí doiciméid a ghiniúint le haghaidh torthaí ar chuardaigh (achoimriú asbhainteach).
- Tá gá le taighde GTN chun córais a fhorbairt inar féidir achoimriú uathoibríoch a dhéanamh ar théacs i gcomhthéacs digiteach (.i. alt nuachta ar líne) i riocht atá níos fusa nó níos sciobtha le léamh (achoimriú coimritheach).
- D'fhonn gur féidir le pobal labhartha na Gaeilge leas a bhaint as an úsáid níos minice a bhaintear as botaí comhrá i seirbhísí ar líne, ní mór tabhairt faoi thaighde leitheadach ar idirghníomhaíochtaí simplí agus lena chois sin taighde leitheadach ar chórais ardchaighdeáin atá feasach ar mhothúcháin.
- Beidh córais Chliste um Fhoghlaim Ríomhchuidithe Teangacha (Caibidil 14) i dtuilleamaí uirlisí iontaofa GTN ar ardchaighdeán a bheith ar fáil. Tá taighde crosdisciplíneach de dhíth, dá bhrí sin.
- Bíonn córais chomhrá (Caibidil 10) i dtuilleamaí córais GTN chun urlabhra atá ar chaighdeán an duine a sholáthar (i bhformáid téacs ar dtús).

3 https://openai.com/api/

Tagairtí

[1]: Reiter, E. and Dale, R., 2000. Building natural language generation systems. Cambridge university press.

[2]: Albert Gatt and Emiel Krahmer. 2018. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. Journal of Artificial Intelligence Research, 61:65–170.

[3] Dale, R., Scott, D. and Di Eugenio, B., 1998. Introduction to the special issue on natural language generation. Computational Linguistics, 24(3), pp.346-353.

[4]: Novikova, J., Dušek, O. and Rieser, V., 2017. The E2E dataset: New challenges for end-to-end generation. arXiv preprint arXiv:1706.09254.

[5]: Gardent, C., Shimorina, A., Narayan, S. and Perez-Beltrachini, L., 2017. The webnlg challenge: Generating text from rdf data. In Proceedings of the 10th International Conference on Natural Language Generation (pp. 124-133).

[6]: Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M. and Schneider, N., 2013. Abstract meaning representation for sembanking. In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse (pp. 178-186).

[7]Goldberg E, Driedger N, Kittredge R (1994). "Using Natural-Language Processing to Produce Weather Forecasts". IEEE Expert. 9 (2): 45–53. doi:10.1109/64.294135.

[8] Reiter E, Sripada S, Hunter J, Yu J, Davy I (2005). "Choosing Words in Computer-Generated Weather Forecasts". Artificial Intelligence. 167: 137–69. doi:10.1016/j.artint.2005.06.006.

[9] Belz A (2008). "Automatic Generation of Weather Forecast Texts Using Comprehensive Probabilistic Generation-Space Models". Natural Language Engineering. 14: 431–55. doi:10.1017/s1351324907004664.

[10] Sripada, N Burnett, R Turner, J Mastin, D Evans(2014). Generating A Case Study: NLG meeting Weather Industry Demand for Quality and Quantity of Textual Weather Forecasts. Proceedings of INLG 2014

[11] Gatt, A., & Reiter, E. (2009, March). SimpleNLG: A realisation engine for practical applications. In Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009) (pp. 90-93).

Caibidil 7
Samhlacha Urlabhra don Ghaeilge

7.1 7.1 Cad é atá i gceist le samhail urlabhra?

Is cuntas simplithe ar phróiseas í an tsamhail, a léiríonn na príomhthréithe a bhaineann leis an phróiseas sin. Sa chás seo is macasamhail ar phróisis na hurlabhra i bhfoirm mhatamaiticiúil, choincréiteach atá i gceist – ríthábhachtach chun ár saineolas ar struchtúr na Gaeilge labhartha (§2.4) a chur i bhfeidhm sa teicneolaíocht. Is ar bhunús eolais (feasacht teangeolaíochta agus urlabhra) a thógtar an tsamhail agus ní hionann agus samhail a thógtar le teicnící meaisínfhoghlama, atá in úsáid fosta go forleathan i dteicneolaíocht an lae inniu. (Tá plé ar na modhanna éagsúla sin in §1.7.)

Is ar thaifead na cainte féin seachas an teanga scríofa a thógtar samhlacha urlabhra. Is fiú a lua gur trí anailís ar théacsanna a thógtar na hacmhainní i gCaibidil 4 – 6 agus léiríonn siad gnéithe de struchtúir na teanga a fheictear sa téacs. Tá plé ar na difríochtaí idir teanga labhartha agus teanga scríofa in §1.6 agus in §2.1. Léiríonn an tsamhail urlabhra gnéithe de struchtúr na teanga labhartha – agus ina measc, na gnéithe faisnéise nach ndéantar nótáil orthu sa bhfoirm scríofa. Ní samhlacha teibí atá i gceist anseo, ach samhlacha coincréiteacha a léiríonn próisis na hurlabhra daonna, mar atá mínithe i gCaibidil 2.4.

7.2 Cén tábhacht atá leo agus cé dó iad?

Tá na samhlacha bunaithe ar thuiscint innealtóireachta ar urlabhra agus ar fheasacht teangeolaíochta ar struchtúr na cainte i dteanga faoi leith. Mínítear na buntáistí a bhaineann leis an nasc taighde seo idir innealtóireacht agus teangeolaíocht, ar bhealach ginearálta, in §7.2.1. Pléitear ansin tábhacht samhlacha ar leith a bhaineann leis an ghlór/prosóid na cainte (§7.2.2) agus le gné na hurlabhraíochta (§7.2.3) agus tugtar samplaí de na himpleachtaí do bhunteicneolaíochtaí agus d'áiseanna don Ghaeilge.

7.2.1 An nasc idir innealtóireacht agus teangeolaíocht

Déanann an tsamhail, atá bunaithe ar innealtóireacht na hurlabhra agus ar shaineolas teangeolaíochta:

Droichead idir foghraíocht teangeolaíochta > teicneolaíocht na hurlabhra
Aistríonn an tsamhail urlabhra ár saineolas foghraíocht-teangeolaíochta go foirm theicniúil is féidir a úsáid sa teicneolaíocht, agus is féidir a léiriú in áiseanna don phobal. Gan a leithéid seo de dhroichead, fanfaidh mórán den taighde ar an Ghaeilge labhartha (§2.4) sa saol acadúil.

Leathnú ar ár dtuiscint ar an teanga labhartha
Is uirlis chumhachtach anailíse í an tsamhail, a chuireann ar ár gcumas gnéithe den teanga labhartha a léiriú nach féidir teacht orthu leis na gnáthuirlisí anailíse i réimse na teangeolaíochta (tá samplaí sna míreanna a leanas).

Nasc idir cainteoir agus éisteoir
Sa chumarsáid labhartha, bíonn cainteoir agus éisteoir i gceist (féach léaráid 15.1 i gCaibidil 15) agus tá an dá thaobh den chumarsáide ríthábhachtach i dteicneolaíocht na hurlabhra, ina bhfuil an chaint á giniúint agus á haithint. Cuirtear an dá ghné seo san áireamh i dtógáil na samhla. Tógtar an tsamhail sa chéad áit bunaithe ar an aschur ón chainteoir dhaonna. Nuair a bhaintear úsáid as an tsamhail chun caint a ghiniúint le meaisín, is féidir tréithe (i.e. paraiméadair) na samhla a láimhseáil agus a mhodhnú i dtástálacha éisteachta a thugann eolas dúinn ar an bhealach a n-airíonn an t-éisteoir iad. Ar an bhealach seo, is féidir tábhacht na dtréithe seo don éisteoir (agus don teicneolaíocht) a mheas.

Bunteicneolaíocht agus áiseanna níos fónta don Ghaeilge
Tá tábhacht faoi leith le samhlacha urlabhra do theicneolaíochtaí agus d'áiseanna inár mian linn teacht ar ghnéithe eolais atá sa chaint, nach bhfuil fáil orthu sa téacs (féach §7.2.2).

Saineolas a aistriú chuig teagasc na teanga labhartha agus chuig an teiripe chainte
Deirtear gur i dteagasc fhuaimniú na teanga (fuaimeanna is prosóid) is lú a éiríonn le teagasc na Gaeilge. Is minic easpa tuisceana ag an ghnáthdhuine, ag múinteoirí agus ag foghlaimeoirí ar ghnéithe foghraíocht-teangeolaíochta – i.e. patrúin struchtúracha na teanga labhartha – i

gcomparáid lena dtuiscint ar ghnéithe gramadaí, atá le feiceáil sa teanga scríofa. Má chuirtear leagan simplithe de shamhail i láthair ar bhealach físiúil, is féidir léiriú coincréiteach a thabhairt ar ghnéithe teibí cainte, ar deacair iad a mhíniú. Tá géarghá lena leithéid i dteagasc na teanga labhartha agus sa teiripe chainte (tá samplaí thíos).

Sna míreanna a leanas, pléitear na samhlacha urlabhra a bhaineann leis an dá phróiseas cainte a bhfuil cur síos orthu i gCaibidil 2.4. Sa chéad áit, le samhail den ghlórfhoinse (foinse na fuaime) [1, 2], is féidir tuiscint a fháil ar ghnéithe iomlána na prosóide, i.e., na hathruithe leanúnacha a dhéantar ar na foinn, ar cháilíocht an ghutha, ar aiceann agus ar rithim na cainte [3-6]. Sa dara háit, tugann samhail urlabhraíochta [7-9] eolas ar ghluaiseacht na mball urlabhra agus ar chruth an bhéil, áit a ndéantar scagadh ar thonn fuaime an ghlóir, agus idirdhealú ar na consain agus na gutaí éagsúla sa teanga. Beidh an dá chineál samhla tábhachtach do theicneolaíocht agus d'áiseanna don Ghaeilge amach anseo.

7.2.2 Samhail de phrosóid an ghlóir: beocht agus brí na cainte

Luíonn cuid mhaith de bhrí na cainte sa phrosóid – sa dóigh a ndeirtear na focail, seachas sna focail féin (féach §2.1). Is fíorbheagán den phrosóid a fheictear sa teanga scríofa. Níl spás anseo le himpleachtaí iomlána na prosóide do bhrí na cainte a phlé, agus mar sin, dírítear ar ghné thábhachtach amháin – faoi mar a chuireann muid ár mothúcháin agus ár meon i leith an éisteora in iúl, trí mhodhnú ar cháilíochtaí an ghutha.

I gcórais sintéise an lae inniu, fuaimníonn an glór an-nádúrtha agus tá siad oiriúnach chun téacs (neodrach) a léamh amach. Ar an drochuair ní féidir leo mothúcháin nó beocht a léiriú sa chaint, agus go minic is air seo a bhraitheann ár dtuiscint iomlán ar bhrí na cainte. Amach anseo, beidh córais chomhrá (Caibidil 10) in úsáid go forleathan, i.e., córais a "éisteann" linn is a "fhreagraíonn" muid, agus a dhéanann aithris ar chomhrá daonna. Ina leithéid seo de chóras, gan an malartú prosóide cuí, fuaimníonn an chaint leadránach, marbhánta agus mí-oiriúnach.

Beidh na córais seo in úsáid sa ghnáthshaol (Caibidil 13) agus tá féidearthachtaí móra ag baint leo do theagasc na teanga, m.sh., i gcluichí idirghníomhacha inar féidir le foghlaimeoirí comhrá a dhéanamh le carachtair fhíorúla a éisteann agus a labhraíonn leo (Caibidil 14). Ina leithéid d'áis, tá gá le glórtha atá in ann aithris a dhéanamh ar an bheocht agus ar na mothúcháin a bhíonn i gcomhrá daonna. Tá sé seo ríthábhachtach fosta i gcás daoine gan urlabhra, ina bhfeidhmíonn an glór sintéiseach mar a nglór féin: tugann an teicneolaíocht glór dóibh, ach ní cheadaíonn sé dóibh an glór sin a mhodhnú mar is mian leo chun iad féin a chur in iúl.

Níl tuiscint mhaith ar an ghné seo den phrosóid go fóill. Tá neamart déanta ag teangeolaithe uirthi mar go bhfuil sé ródheacair cáilíochtaí an ghlóir a scrúdú le gnáth-uirlisí teangeolaíochta. Le samhail den ghlór, is féidir anailís a dhéanamh ar an tslí a léiríonn modhnú ar cháilíochtaí an ghlóir mothúcháin is meon an chainteora [10,11]. Is féidir an tsamhail chéanna a úsáid sa teicneolaíocht, m.sh sa tsintéis chun cáilíochtaí an ghutha a smachtú agus beocht is mothúcháin a chur in iúl [12, 13].

Tá impleachtaí leathana ag an ghné seo cainte, ní hamháin do shintéis na cainte, ach do na teicneolaíochtaí uile a bhaineann leis an teanga labhartha. Faoi láthair, ní aithníonn córais aitheanta cainte (Caibidil 9) ach na focail a deirtear (mar atá sa leagan scríofa) agus níl teacht acu ar an eolas idirphearsanta a bhaineann le meon agus le mothúcháin. Táthar ag díriú anois ar an bhearna eolais seo, agus tá forbairt á dhéanamh ar áiseanna a aimsíonn na gnéithe seo den chumarsáid dhaonna – m.sh., in ionaid glaonna a bhaineann úsáid as an aithint cainte agus in áiseanna sláinte. Léiríonn prosóid an ghlóir eolas ar shláinte agus ar mheabhairshláinte duine, agus tá áiseanna á bhforbairt lena ndéantar monatóireacht ar cháilíochtaí gutha le gur féidir le lucht leighis tacaíocht thráthúil a chur ar fáil. I gcás córais comhrá (Caibidil 10), ní hamháin go mbeifear ag dréim le meaisín a labhraíonn leat le mothúcháin, ach fosta a "aithníonn" na mothúcháin i do chuid cainte.

Don fhoghlaimeoir, tá dúshlán ar leith ag baint le sealbhú na prosóide dúchasaí. Tá an ghné seo den teanga fíordheacair a theagasc, go háirithe mar nach bhfeictear an phrosóid sa teanga scríofa. Ach nuair is féidir taispeántas físiúil den tsamhail a léiriú, beifear in ann díriú ar an ghné thábhachtach seo den teanga labhartha. Is féidir an tsamhail a chur in oiriúint do phatrúin phrosóide na gcanúintí éagsúla (§2.4). Bheadh taispeántas físiúil den tsamhail thar a bheith úsáideach don teiripe chainte chomh maith.

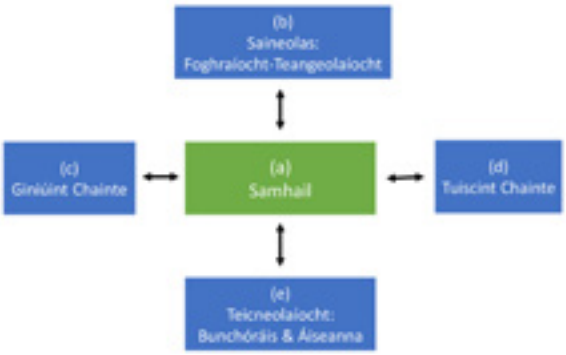
7.2.3 Samhlacha urlabhraíochta

Léiríonn samhail urlabhraíochta gluaiseacht na mball urlabhra agus an nasc le tréithe acústaice na bhfuaimeanna. Tá tábhachtach ar leith leis seo i gcás na Gaeilge, mar gheall ar an idirdhealú caol/leathan i measc chonsain na Gaeilge, atá an-difriúil leis an Bhéarla. Mar shampla, tá fuaimniú iomlán difriúil ag an chéad chonsan sa dá fhocal, /lʸoːnʸː/ 'lón' agus /lʲoːnʸː/ 'leon', agus ní hionann ceachtar acu agus an consan ag tús an fhocail /loːn/ 'lone' sa Bhéarla (féach §2.4).

Go dtí le déanaí, níorbh fhéidir samhlacha críochnúla urlabhraíochta a thógáil a bheadh oiriúnach don teicneolaíocht, mar nach raibh ar ár gcumas taifead ar ardchaighdeán agus ar ardluas a fháil de ghluaiseacht na mball urlabhra. Ach le blianta beaga anuas, leis an dul chun cinn le córais fMRI, tá borradh tapa ag teacht sa réimse taighde seo [14, 15]. Amach anseo beidh féidearthachtaí ann d'áiseanna nuálacha a bheas bunaithe ar shamhlacha urlabhra, m.sh. innill urlabhraíochta cainte don tsintéis [9] nó córais aitheanta cainte ina mbaintear úsáid as tréithe urlabhraíochta.

Beidh samhlacha urlabhraíochta úsáideach fosta don teagasc agus do theiripe na cainte. Is minic a bhíonn deacracht ag foghlaimeoirí le hurlabhraíocht na gconsan Gaeilge. Gan sealbhú ceart ar idirdhealú caol/leathan na gconsan, ní féidir blas dúchasach a bhaint amach, agus cruthaíonn sé deacrachtaí do shealbhú na litearthachta mar aon le gnéithe áirithe gramadaí (§2.4.2 is §2.7.3). Is féidir tairbhe a bhaint as samhlacha urlabhraíochta le haiseolas coincréiteach, físiúil, chun fuaimeanna dúchasacha teanga a theagasc [16], rud a bheadh an-úsáideach i gcás na Gaeilge. Mar an gcéanna don teiripe urlabhra: dóibh siúd a bhfuil fadhbanna urlabhraíochta acu, d'fhéadfaí, de thoradh na samhla, soiléiriú agus oiliúint a thabhairt ar thréithe den fhuaimniú ar deacair iad a mhíniú agus nach bhfuil le feiceáil go soiléir sa scríobh.

7.3 Cén dóigh a n-oibríonn an tsamhail?



Fíor 7.1 Samhail urlabhra mar mhol lenar féidir feasacht foghraíocht-teangeolaíochta a chur i bhfeidhm i dteicneolaíocht na hurlabhra, trí phróiseas leanúnach forbartha

Cruthaítear an tsamhail urlabhra (a) sa chéad áit le sonraí a eascraíonn as saineolas foghraíocht-teangeolaíochta (b) (Caibidil 2.4). Le tástáil a dhéanamh ar an tsamhail, déantar comparáid idir aschur na samhla agus urlabhra an chainteora dhaonna (c) agus déantar trialacha chun aireachtáil an éisteora dhaonna (d) a mheas freisin. Le torthaí na dtástálacha sin bíonn an deis ann an tsamhail a fheabhsú agus tagann méadú ar ár dtuiscint ar struchtúr na theanga. Trí shamhail a chur i bhfeidhm in áis teicneolaíochta (d), cruthaítear ardán le haiseolas a bhailiú ón lucht úsáide. Faightear léargas ar bhuanna agus ar laigí na samhla, rud a chuireann ar ár gcumas tuilleadh leasúchán a dhéanamh ar an tsamhail, agus dá bharr sin cuirtear lenár n-eolas foghraíochta, agus le dul chun cinn na teicneolaíochta.

7.4 A bhfuil déanta i leith na Gaeilge go dtí seo

Samhail de Phrosóid an Ghlóir: Tá samhail pharaiméadrach den ghlór [1, 2] in úsáid chun anailís a dhéanamh ar ghnéithe de phrosóid na Gaeilge a bhaineann le cáilíochtaí an ghutha [5, 13, 17] sa tionscadal Róbóglór i gColáiste na Tríonóide. Tá an taighde sin ag tógáil ar anailís ghlóir agus ar thrialacha aireachtála le héisteoirí [6, 10, 11]. Cuireann sé fosta leis an bhuntaighde teangeolaíochta a rinneadh ar thuiníocht na gcanúintí (luaite in §2.4.3): amach anseo, táthar ag súil go mbeidh léargas níos iomláine le fáil ar phrosóid na Gaeilge, ina mbeidh modhnú an ghutha le mothúcháin san áireamh. Bunaithe ar an obair sin tá bealaí á bhfiosrú le tréithe an ghlóir a smachtú sa tsintéis, agus tá trialacha ar bun chun torthaí an taighde a chur i bhfeidhm i nglórtha sintéiseacha ABAIR [13].

Samhail Urlabhraíochta: Go dtí seo, cé go bhfuil roinnt (bheag) anailís urlabhraíochta déanta ar na consain Ghaeilge (§2.4), mar a mhínítear thuas, ní raibh fáil go dtí le déanaí ar na modhanna taifeadta agus anailíse a mbeadh gá leo le haghaidh samhlacha teicneolaíochta.

7.5 A bhfuil le déanamh

Trasnaíonn an taighde ar shamhlacha urlabhra réimse na foghraíocht-teangeolaíochta agus forbairt na teicneolaíochta agus, mar sin, is gá tabhairt faoi ar bhealach idirdhisciplíneach. Ar thaobh amháin, tá dlúth-chomhoibriú de dhíth le taighdeoirí foghraíochta (§2.4), sa dóigh is gur féidir samhlacha a úsáid mar uirlisí anailíse a thacaíonn le bailiú sonraí, agus go bhfuil na sonraí á n-úsáid le samhlacha a thástáil agus a fheabhsú. Ar an taobh eile, tá sé tábhachtach go ndéantar forbairt ar na samhlacha lámh ar láimh le tógáil na mbunteicneolaíochtaí agus

na n-áiseanna: braitheann foirmliú agus cur i bhfeidhm na samhla ar riachtanais córais nó áise ar leith. Mar shampla, agus córais shintéiseacha á dtógáil (Caibidil 8), is féidir innill labhartha éagsúla a roghnú: le hinnill áirithe d'fhéadfaí an tsamhail féin a úsáid; le cinn eile, bheadh modh indíreach de dhíth chun aschur ón tsamhail a úsáid.

Mórsprioc amháin do shamhail ar phrosóid an ghlóir, ná an guth sintéiseach a smachtú le beocht agus mothúcháin a léiriú (féach Caibidil 8). Cé go bhfuil píosa fada le dul go mbeidh an glór sintéiseach in ann aithris a dhéanamh ar shaibhreas dochreidte na prosóide daonna, is féidir dul chun cinn a dhéanamh céim ar chéim, agus toradh na hoibre a bhrath laistigh de chórais shrianta i bhfeidhmchláir ar leith, m.sh. chun cluiche idirghníomhach teanga nó chun áis rochtana a dhéanamh níos oiriúnaí agus níos tarraingtí d'úsáideoirí. Le himeacht aimsire, ba cheart go dtiocfadh méadú ar na teicneolaíochtaí agus ar réimse na n-áiseanna ina mbeidh teicneolaíocht na hurlabhra in úsáid go héifeachtach don Ghaeilge.

Leis an dul chun cinn atá déanta in úsáid na ngléasanna fMRI le sonraí urlabhraíochta a bhailiú, tá sé indéanta anois tabhairt faoi shamhail faoi choinne urlabhraíocht na Gaeilge. Chun sonraí ar chonsain agus ar ghutaí na Gaeilge a bhailiú, moltar sa chéad áit tabhairt faoi i gcomhpháirt leis na Saotharlanna ceannródaíocha sa réimse seo thar lear. Amach anseo, má tá muid le freastal ar na difríochtaí canúnacha urlabhraíochta agus ar áiseanna don phobal, ba mhaith na modhanna taighde seo a fhorbairt in Éirinn fosta.

Cuirfidh samhlacha críochnúla go mór lenár dtuiscint ar an teanga, agus beifear in ann lántairbhe a bhaint as teicneolaíochtaí nuálacha a bheas ag teacht ar a bhfód, amhail innill labhartha, córais aitheanta cainte agus córais chomhrá ina bhfuil gnéithe prosóide agus urlabhraíochta san áireamh. In áiseanna teagaisc agus sa teiripe urlabhra, cuideoidh na samhlacha linn déileáil le gnéithe tábhachtacha den teanga labhartha nárbh fhéidir díriú orthu go héasca go dtí seo.

Cé gur taighde taiscéalaíoch atá i gceist le dearadh samhlacha, a bhraitheann ar mhodhanna úrscothacha taighde, moltar iad a chur in úsáid ón tús, laistigh d'áiseanna ábhartha – na háiseanna is mó a bhfuil gá leo i saol na Gaeilge. A luaithe a bheas leagan píolótach cruthaithe, is féidir é a úsáid chun measúnú agus leasú a dhéanamh ar an tsamhail. Sa dóigh seo, beidh buntáistí na hoibre chun tairbhe phobal na Gaeilge, fiú más ar bhonn fadtéarmach a fheicfear an tionchar iomlán.

7.6 Moltaí

Mar atá ráite, moltar dlúth-chomhoibriú le taighdeoirí foghraíocht-teangeolaíochta, chun sonraí a bhailiú agus tástáil leanúnach a dhéanamh ar na samhlacha. Mar an gcéanna, beidh comhoibriú de dhíth le lucht tógála na gcóras teicneolaíochta, chun a chinntiú go sroicheann aschur an taighde an pobal agus go bhfuil aiseolas ón phobal ag cur le leasú na samhlacha. Mar sin, tá moltaí don obair seo ceangailte leis na moltaí do na caibidlí a bhaineann le buntaighde ar an teanga labhartha (§ 2.4), le teicneolaíochtaí na hurlabhra (Caibidlí 8, 9 & 10) agus leis na háiseanna atá bunaithe orthu.

Beidh gá le foireann idirdhisciplíneach atá dírithe ar thaighde na Gaeilge, ina bhfuil teangeolaithe (a bhfuil scileanna foghraíochta acu), saineolaithe i dtógáil samhlacha urlabhra agus saineolaithe i dtógáil na dteicneolaíochtaí nuálacha. Mar a mholtar i gCaibidlí 1 agus 16, beidh a leithéid sin de bhonneagar daonna ina bhunchloch don earnáil amach anseo.

1. Samhail a thógáil de phrosóid an ghlóir sna canúintí Gaeilge, chun léiriú a thabhairt ar na príomhthréithe prosóide a bhaineann le brí na cainte labhartha, agus béim faoi leith ar na gnéithe sin atá bainteach le meon agus mothúcháin an chainteora. Beidh gá le dlúth-chomhoibriú idir saineolaithe foghraíochta agus innealtóireachta.

2. Samhail a thógáil d'urlabhraíocht na Gaeilge, ag díriú ar na consain agus a dtionchar ar urlabhraíocht na ngutaí. Beidh ionchur de dhíth ó shaineolaithe i saotharlanna thar lear don obair seo, chomh maith le comhoibriú idir saineolaithe foghraíochta agus innealltóireachta.

3. Taighde a dhéanamh ar bhealaí leis na samhlacha seo a chur in úsáid sna bunteicneolaíochtaí – is é sin, sna córais sintéise agus aitheanta cainte. Beidh cur i bhfeidhm na samhlacha sna córais seo ag brath ar na hinnill labhartha agus chainte atá in úsáid sna córais, agus mar sin, beidh comhoibriú i measc na ngrúpaí taighde de dhíth.

4. Na samhlacha ag aschur as 1 & 2 a chur in úsáid laistigh d'áiseanna – mar chruthú coincheapa sa chéad áit. Moltar díriú go speisialta ar áiseanna do theagasc na Gaeilge, do theiripe urlabhra agus dóibh siúd faoi mhíchumas. Moltar sa dá réimse (prosóid an ghlóir agus an urlabhraíocht) go ndéanfaí trialacha le leaganacha simplithe ar scáileáin. Bheadh sé tábhachtach go ndéanfaí an taighde seo i gcomhpháirt le húsáideoirí, a gcuid múinteoirí, lucht teiripe srl.

Tagairtí

[1] Fant, G., Liljencrants, J. and Lin, Q. (1985). A four-parameter model of glottal flow. STL-QPSR, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 26 (4), pp. 1-13.

[2] Fant, G. (1997). The voice source in connected speech. Speech Communication, 22, 125-139.

[3] Gobl, C., & Ní Chasaide, A. (2010). Voice source variation and its communicative functions. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), The Handbook of Phonetic Sciences (2 ed., pp. 378-423). Oxford: Blackwell Publishing Ltd.

[4] Iseli, M., Shue, Y.-L., Epstein, M. A., Keating, P., Kreiman, J., and Alwan, A. (2006). Voice source correlates of prosodic features in American English: A Pilot Study. Proceedings of Interspeech 2006, Pittsburgh, PA, pp. 2226-2229.Iseli

[5] Ní Chasaide, A., Yanushevskaya, I. and Gobl, C. (2015). Prosody of voice: declination, sentence mode and interaction with prominence. Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK

[6] Yanushevskaya, I., Gobl, C. and Ní Chasaide, A. (2017). Cross-speaker variation in voice source correlates of focus and deaccentuation. INTERSPEECH 2017, Stockholm, Sweden, pp. 1034-1038.

[7] Mermelstein, P. (1973). Articulatory model for the study of speech production. Journal of the Acoustical Society of America 53: 1070–1082.

[8] Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In: Hardcastle WJ, Marchal A, editors, Speech production and speech modelling, Kluwer Academic Publishers, Boston. 131–149.

[9] Birkholz, P. (2013). Modeling consonant-vowel coarticulation for articulatory speech synthesis. PloS One, 8(4).

[10] Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. Speech Communication, 40(1-2), 189-212.

[11] Yanushevskaya, I., Gobl, C., & Ní Chasaide, A. (2018). Cross-language differences in how voice quality and f0 contours map to affect. Journal of the Acoustical Society of America, 144(5), 2730-2750.

[12] Sorin, A., Shechtman, S. and Rendel, A. (2017). Semi Parametric Concatenative TTS with Instant Voice Modification Capabilities. Proceedings of INTERSPEECH 2017, Stockholm Sweden, pp. 1373–1377.

[13] Murphy, A., Yanushevskaya, I., Ní Chasaide, A. and Gobl, C. (2021). Integrating a Voice Analysis-Synthesis System with a TTS Framework for Controlling Affect and Speaker Identity. Proc. 32nd Irish Signals and Systems Conference (ISSC 2021), Athlone, Ireland.

[14] Narayanan, S., Toutios, A., Ramanarayanan, V., Lammert, A., Kim, J., Lee, S., Proctor, M. (2014). Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC). The Journal of the Acoustical Society of America, 136(3), 1307–1311.

[15] Ramanarayanan, V., Tilsen, S., Proctor, M., Töger, J., Goldstein, L., Nayak, K. S., & Narayanan, S. (2018). Analysis of speech production real-time MRI. Computer Speech & Language, 52, 1–22.

[16] Suemitsu, A., Dang, J., Ito, T., & Tiede, M. (2015). A real-time articulatory visual feedback approach with target presentation for second language pronunciation learning. The Journal of the Acoustical Society of America, 138(4).

[17] Murphy, A., Yanushevskaya, I., Ní Chasaide, A. and Gobl, C. (2018). Voice Source Contribution to Prominence Perception: Rd Implementation. INTERSPEECH 2018, Hyderabad, India, pp. 217-221.

[18] Ní Chasaide, A., Yanushevskaya, I., Kane, J., Gobl, C. (2013). The voice prominence hypothesis: the interplay of F0 and voice source features in accentuation. INTERSPEECH 2013, Lyon, France, 3527-3531.

Caibidil 8
Sintéis Chainte

8.1 Cad is sintéis chainte ann?

Is caint a ghineann meaisín í an tsintéis chainte (SC). De ghnáth, tugtar téacs don chóras agus léitear amach os ard é – mar sin, aistríonn sé téacs-go-caint. Cluintear guthanna sintéise go forleathan, m.sh. ón ríomhaire, ón chúntóir phearsanta ar do ghuthán, san ardaitheoir, i gcluichí ríomhaire, agus is glór é ag mórán daoine faoi mhíchumas. Is féidir leis an ghlór sintéise a bheith nádúrtha mar a bheadh guth daonna ann, nó mínádúrtha, mar a bhí sa chóras a d'úsáid Stephen Hawking.

Mar nach bhfuil aon chanúint amháin mar chaighdeán labhartha sa Ghaeilge, is gá freastal ar na canúintí éagsúla – rud atá fíor do theicneolaíochtaí uile na hurlabhra don teanga.

8.2 Cén tábhacht atá léis agus cé dó é?

Tá an tsintéis ríthábhachtach do theangacha atá i mbaol, mar go gcuirtear an teanga labhartha i lár an aonaigh i saol digiteach an lae inniu. Más féidir an Ghaeilge labhartha a chur ar fáil sna háiseanna a mbaineann muid úsáid laethúil astu, spreagfaidh sé úsáid na teanga i réimsí níos leithne dár ngnáthshaol.

Bíonn tábhacht na sintéise agus an úsáid a bhaintear aisti difriúil in amanna sa mhionteanga i gcomparáid le teanga fhorleathan mar an Béarla. Is féidir an tábhacht don Ghaeilge a léiriú ó shamplaí d'úsáid na nglórtha sintéise atá forbartha cheana féin do chanúintí na Gaeilge i dtionscadal ABAIR i gColáiste na Tríonóide (tá cur síos ar áiseanna ar leith i mír §8.4).

Áis do phobal forleathan na Gaeilge, na himircigh san áireamh: tá an-éileamh ar shuíomh gréasáin ABAIR ina bhfuil na glórtha ar fáil do na mórchanúintí. Is ón choigríoch a lán de na cuairteoirí, rud a léiríonn chomh scaipthe is atá líonraí pobail na Gaeilge. Feictear san aiseolas ón lucht úsáide gur mór an cuidiú an chaint bheo dhúchasach a bheith ar fáil don teanga scríofa (tá tuilleadh plé air seo thíos).

Áiseanna rochtana dóibh siúd le míchumas: is áis bhunúsach an tsintéis do mhórán daoine faoi mhíchumas. Dóibh siúd faoi mhíchumas radhairc, gan léitheoir scáileáin le sintéis chainte, tá sé beagnach dodhéanta páirt a ghlacadh san oideachas trí Ghaeilge nó i saol sóisialta agus cultúrtha na Gaeilge ar líne. Ceadaíonn an tsintéis áiseanna bunúsacha cumarsáide is rochtana dóibh siúd atá gan urlabhra (de bharr uathachas, pairilis cheirbreach, galar Parkinson srl.). Is áis thábhachtach an tsintéis chainte do mhórán le disléicse. Pléitear na gnéithe seo i gCaibidil 15.

Oideachas: tá impleachtaí speisialta don mhionteanga i réimse an oideachais. Ní chluineann formhór fhoghlaimeoirí Gaeilge an chaint dhúchasach go hiondúil, ach le córas sintéise is féidir 'cainteoir dúchais fíorúil' a bheith agat sa seomra ranga agus sa bhaile. Nuair a chluineann tú an chaint bheo in éineacht leis an fhocal scríofa, cuidíonn sé le fuaimeanna na teanga a fháil ar do chluas agus scileanna labhartha agus éisteachta a shealbhú. Cothaítear tuiscint ar an chomhfhreagairt idir an fhuaim agus litreacha scríofa – atá sách doiléir sa Ghaeilge – rud a thacaíonn le sealbhú na litearthachta (féach §2.4.2 & §2.7.3). De thoradh na sintéise tá deis ann forbairt a dhéanamh ar áiseanna cumhachta teagaisc mar ardáin idirghníomhacha agus ríomhchluichí oideachasúla – agus féidearthachtaí móra ann do theagasc na teanga agus do spreagadh na bhfoghlaimeoirí (féach Caibidil 14).

Caomhnú agus doiciméadú ar an teanga: mar chainteoir dúchais fíorúil, déanann an glór sintéise caomhnú agus doiciméadú ar chanúint faoi leith [1, 2]. I gcás canúint atá i mbaol, is luachmhar an áis 'cainteoir fíorúil' do phobal atá ag iarraidh a chanúint a athbheochan nó a chaomhnú, go háirithe nuair a chuirtear san áireamh na hacmhainní teanga a ndéantar forbairt orthu i dtógáil na sintéise (§8.3.1) agus na háiseanna is féidir a bhunú leis (§8.4.2). Fiú i gcás canúint atá caillte, má tá taifeadtaí oiriúnacha ar fáil, d'fhéadfaí 'aiséirí' a shamhlú i bhfoirm cainteoir fíorúil.

Taighde acadúil ar chanúintí na Gaeilge: sa mhéid is go ndéantar forbairt ar acmhainní teangeolaíochta (§2.4), ar chorpais (Caibidil 3) agus ar shamhlacha urlabhra (Caibidil 7) cuirtear lenár dtuiscint ar chanúintí na Gaeilge agus cuirtear acmhainní luachmhara agus

modhanna taighde úra ar fáil do thaighdeoirí na todhchaí.

Nuair a nascfar an tsintéis le córais eile, mar an aithint cainte (Caibidil 9) i gcórais chomhrá idirghníomhacha (Caibidil 10), tiocfaidh méadú mór ar réimse na n-áiseanna agus ar impleachtaí na teicneolaíochta. Amach anseo, is léir go mbeidh muid ag caint is ag comhrá le gléasanna leictreonacha agus le róbait. Tiocfaidh áiseanna eile chun tosaigh, m.sh an ateangaireacht uathoibríoch ina nasctar teicneolaíochtaí na hurlabhra le córais mheaisínaistriúcháin.

8.3 Cén dóigh a n-oibríonn sé?

Tá modhanna éagsúla le córas sintéise a thógáil [3]. Tá dhá mhórghrúpa i gceist (a) córais mhodúlacha, ina bhfuil corpais labhartha, acmhainní teangeolaíochta-urlabhra agus inneall labhartha san áireamh, agus (b) córais Ceann go Ceann (CgC) nach mbaineann (mórán) úsáide as acmhainní teangeolaíochta, ach a bhaineann úsáid as modhanna domhainfhoghlama chun an chomhfhreagairt idir an taifead cainte agus an téacs a 'fhoghlaim'. Baineann an modh seo ardchaighdeán sintéise amach i dteangacha mar an Béarla. Mar sin féin, tá na córais CgC ag brath ar chorpais ollmhóra agus ar ríomhairí sárchumhachtacha.

8.3.1 Córais mhodúlacha

Léirítear i bhFíor 8.1 na trí mhodúl a ndéantar forbairt orthu sna córais seo. Tá an modúl teangeolaíochta-urlabhra mar a bheadh droichead idir an corpas cainte agus an t-inneall labhartha.

Na Corpais

Sa mhodh seo, tógtar an glór sintéise (do theanga nó do chanúint ar leith) mar mhacasamhail cainteora dúchais amháin. Déantar taifead ar ardchaighdeán ar chorpas (dreas fada cainte) den chainteoir atá roghnaithe. Déantar tras-scríobh foghraíochta air agus ailíniú ar gach consan agus guta leis an chomhad fuaime. Tá dearadh an chorpais tábhachtach: bíonn an-tionchar ag an chomhthéacs ar thréithe na fuaime i.e., na fuaimeanna roimhe agus ina diaidh, agus an comhthéacs prosóide. Mar sin, is gá go mbíonn clúdach sa chorpas ar gach fuaim i ngach comhthéacs ina bhfaightear í. Freagraíonn corpas ARTIC [4] do na riachtanais seo sa Bhéarla. Mar gheall ar líon mór na gconsan Gaeilge (§2.4.2) tá corpas i bhfad níos mó de dhíth. Caithfidh an corpas freagairt do chóras fuaime na gcanúintí éagsúla, agus is ceart don ábhar a ndéantar taifead air a bheith in oiriúint don chanúint.

Fíor 8.1: Sa chóras modúlach sintéise, déantar forbairt ar chorpais chainte, acmhainní teangeolaíochta-urlabhra, agus inneall labhartha.

Acmhainní Teangeolaíochta agus Urlabhra

Baineann an córas sintéise úsáid as acmhainní teangeolaíochta agus samhlacha urlabhra don chanúint faoi leith (Caibidil §2.4 & 7) a cheadaíonn téacs agus siombailí a aistriú go sraith fuaimeanna le gnéithe prosóide, agus a ullmhaíonn an t-ionchur don inneall labhartha. Sa chéad áit, déantar normalú ar an téacs (athraítear uimhreacha, giorrúcháin, siombailí srl. go téacs). Ansin aistríonn na rialacha litir-go-fuaim (§2.4.4) an téacs isteach ina shraith fuaimeanna (m.sh. buíon /bɤi:nɤ/). I gcás eisceachtaí ar na rialacha seo, ina bhfuil litriú neamhrialta i gceist, baintear úsáid as foclóir fuaimnithe don chanúint – m.sh., (ní) bhfaighfidh /wɤ:i/ i dTír Chonaill agus /vɤaɪgʲ/ i gCiarraí (féach §2.7.3). Gintear samhlacha acústaice le heolas ar thréithe acústaice i leith gach fuaime i ngach comhthéacs. Tugann samhlacha prosóide eolas ar an na patrúin tuiníochta, aicinn agus ghutha (Caibidil 7 agus §2.4.3).

Braitheann rogha agus dearadh na n-acmhainní teangeolaíochta agus urlabhra ar an inneall labhartha a mbaintear úsáid as. Caithfidh na hacmhainní uile a bheith in oiriúint don chanúint.

An tInneall Labhartha

Tógann an t-inneall labhartha an tsraith fuaimeanna (de ghnáth, na samhlacha acústaice agus tréithe áirithe prosóide san áireamh) agus gineann sé an t-aschur labhartha. Tá saghsanna éagsúla innill labhartha is féidir a úsáid.

Sintéis pharaiméadrach (foinse – scagaire): gintear an tsintéis díreach ó shamhlacha acústaice den ghlór (foinse na fuaime) agus den urlabhraíocht (scagadh na fuaime) (féach Caibidil 7 & §2.4) . Is mar seo a tógadh na chéad chórais sintéise, agus tá a leithéid d'innill in úsáid go fóill, go mór mór in áiseanna dóibh siúd atá faoi mhíchumas. Ag brath ar chríochnúlacht shamhail an ghlóir, is féidir leis an aschur cainte a bheith an-nádúrtha nó ar nós róbait (mar a bhí ag Steven Hawking).

Sintéis atá bunaithe ar roghnúchán aonaid: modh an-difriúil atá i gceist anseo, nach ngineann an chaint ó shamhlacha acústaice. Déantar cuardach i gcorpas ollmhór ar na fuaimeanna (nó ar mhionchodanna fuaime) sa chomhthéacs atá de dhíth, agus snitear le chéile iad gan aon phróiseáil a dhéanamh orthu [5]. Is féidir aschur cainte an-nádúrtha a bhaint amach, ach in amanna 'teipeann air' le drochearráidí, ar fháthanna éagsúla, m.sh. má tá bearnaí i gclúdach na bhfuaimeanna sa chorpas, botúin sa tras-scríobh srl. Mar go bhfuil corpas ollmhór le stóráil agus le cuardach chun aschur cainte a ghiniúint, tá sé rómhall d'áiseanna ina bhfuil ardluas de dhíth (m.sh. léitheoir scáileáin do na daill) agus rómhór do ghléasanna beaga mar an fón póca.

Sintéis pharaiméadrach le meaisínfhoghlaim (SPSS): gintear an chaint ó shamhlacha acústaice (do gach fuaim i ngach comhthéacs) a thógtar trí mhodhanna meaisínfhoghlama (féach Caibidil 1.7), ag baint úsáid as HMManna [6, 7], agus líonraí néaracha domhainfhoghlama DNNanna [8, 9]. Tá na hinnill labhartha seo in úsáid go forleathan, agus is féidir leo aschur cainte ar ardchaighdeán a ghiniúint, atá tapa agus seasmhach.

Sintéis atá bunaithe ar phróiseas na hurlabhraíochta (daonna): gintear an chaint mar mhacasamhail ar phróiseas na hurlabhraíochta daonna, i.e. ó shamhlacha meicniúla den ghlór agus de scagadh na hurlabhra (tá cur síos orthu seo i gCaibidil 7 & §2.4). Cé nach bhfuil an modh sintéise seo ar fáil (go fóill) taobh amuigh de Shaotharlanna Urlabhra, tá féidearthachtaí móra aige don todhchaí – go háirithe mar gur féidir anois corpais lena ngabhann sonraí urlabhraíochta a bhailiú (Caibidil 7). Mar gur samhail de phróisis na hurlabhra daonna atá i gceist, d'fhéadfaí leis an modh seo sintéis chríochnúil sholúbtha a chur ar fáil. Mar shampla, má athraítear toisí na mball urlabhra agus tréithe shamhail an ghutha, d'fhéadfaí glórtha éagsúla (fir, mná agus páistí) a ghiniúint ón aon ghuth sintéise amháin. D'fhéadfaí cáilíochtaí an ghlóir a mhodhnú chun beocht agus mothúchán a léiriú sa ghuth sintéise (féach thíos, agus Caibidil 7).

8.3.2 Córais ó Cheann go Ceann

Tá forbairt an-tapa ag teacht ar theicneolaíocht na sintéise. Tá fás ar úsáid líonraí néaracha domhainfhoghlama chun córais ó Cheann go Ceann (CgC) a thógáil. Sna córais sin, tá gá le corpais labhartha ar scála ollmhór lena ngabhann líon mór cainteoirí, agus ní bhaintear úsáid as na hacmhainní teangeolaíochta ach ar bhealach indíreach [10 - 14]. Oibríonn na córais CgC i dteanga mar an Béarla, mar a bhfuil fáil ar ollchorpais lena ngabhann líon mór cainteoirí, rud nach fíor do na teangacha neamhfhorleathana. Mar sin féin, beidh féidearthachtaí móra ann don Ghaeilge nuair a bheas corpais fhorleathana labhartha ar fáil do na canúintí éagsúla.

In ainneoin na mbuntáistí móra le córais CgC, tá fadhbanna leo agus beidh gá fós le córais ina bhfuil feasacht teangeolaíochta suite (mar atá sna córais mhodúlacha, pléite thuas). Fadhb amháin a ghabhann le córais CgC is ea nach féidir earráidí san aschur cainte a cheartú mura bhfuil teacht againn ar an eolas teangeolaíochta. Tá fadhb theicniúil a bhaineann le méid na mbogearraí agus le cumhacht na ríomhairí a mbíonn gá leo. Ciallaíonn seo, faoi láthair, nach féidir a leithéid de chórais sintéise a lonnú laistigh d'áiseanna a chuirtear ar fáil ar ghléasanna beaga ina bhfuil srian ar an chuimhne, nó is gá a reáchtáil as líne. Níos tábhachtaí fós i gcás na Gaeilge de, tá an fheasacht teangeolaíochta i gcroílár an chórais de dhíth i gcás mórán de na háiseanna a bheadh úsáideach don phobal (féach §8.4 thíos).

Is léir go bhfuil an dá shaghas córas tábhachtach don Ghaeilge, agus go dtiocfaidh córais hibrideacha chun tosaigh ina mbainfear tairbhe as na buntáistí faoi leith a bhaineann le gach ceann acu.

8.4 A bhfuil déanta go dtí seo i leith na Gaeilge

Is réimse é seo ina bhfuil roinnt maith taighde ar siúl. Tá forbairt déanta ar chórais TgC i gcomhair chanúintí na Gaeilge, lámh ar lámh le háiseanna don phobal. Pléitear iad seo sna míreanna a leanas.

8.4.1 Córais téacs-go-caint don Ghaeilge

Tá forbairt déanta ar chórais sintéise, bunaithe ar chainteoirí dúchais ó na trí mhórchanúint, agus tá siad ar fáil saor in aisce ar shuíomh idirlín ABAIR : glór Ultach (Gaoth Dobhair, baineann); glór Chonnachtach (Conamara, fireann); glór Muimhneach (Corca Dhuibhne, baineann agus fireann). Tá modhanna éagsúla sintéise agus innill labhartha in úsáid. Is féidir an chanúint agus an cainteoir a roghnú, chomh maith leis an inneall labhartha (agus an modh sintéise) is fearr a fheileann do riachtanas ar leith. Tá glór fireann agus baineann (glórtha Galltachta) curtha ar fáil le déanaí fosta ag Microsoft, agus tá glór baineann Ultach (Galltachta) ar díol fosta ag comhlacht Albanach, Cereproc.

Tá taighde ar siúl ag ABAIR chun forleathnú a dhéanamh ar na corpais agus ar na glórtha sintéiseacha. Tá taighde

fosta ar siúl ar chórais CgC do na canúintí, agus béim ar chórais hibrideacha ina bhfuil teacht ar mhodúil teangeolaíochta.

8.4.2 Áiseanna atá bunaithe ar shintéis

Ag freagairt don éileamh ó dhaoine agus ó ghrúpaí pobail, tá forbairt déanta ar áiseanna le glórtha sintéiseacha ABAIR. Is féidir teacht ar fhormhór na n-áiseanna a leanas ar shuíomh ABAIR, agus tá cuid eile acu fós á bhforbairt.

Léitheoir ar líne : leis an bhreiseán seo is féidir leathanaigh ghréasáin a léamh amach. Roghnaíonn an t-úsáideoir an chanúint, an t-inneall labhartha agus luas na cainte. Is féidir an téacs a mhéadú agus is féidir aibhsiú a dhéanamh air fosta, de réir mar a léitear amach é.

Aip Android don fhón póca : tá glór Connachtach (le hinneall labhartha HMM) ar fáil faoi láthair, agus tá na canúintí eile á gcur leis.

Léitheoir scáileáin (breiseán NVDA) do dhaoine faoi mhíchumas radhairc : tá breiseán le Gaeilge ar fáil don léitheoir scáileáin ar fhoinse oscailte NVDA, atá saor in aisce. Léitear téacs amach os ard, chomh maith le treoracha chun an ríomhaire agus bogearraí a láimhseáil. Is féidir an córas Braille Liblouis a chur ag obair leis an bhreiseán, chun aschur cainte agus Braille a fháil ag an am chéanna [14] (féach Caibidil 15).

Leabhair Ilmheáin Daisy : Tá na glórtha sintéise in úsáid in Leabhair DAISY – sraith leabhar ilmheáin lena ngabhann aschur cainte do pháistí scoile faoi mhíchumas radhairc.

Córas cumarsáide sintéise (AAC): is córas é seo lenar féidir le daoine gan urlabhra cumarsáid a dhéanamh. Roghnaíonn an t-úsáideoir straith focal/pictiúr le habairt a chumadh agus deirtear amach í le glór sintéiseach. Tá bunchóras don Ghaeilge tógtha ar bhonn trialach [15, 16], ach tá go leor forbartha le déanamh air go fóill (féach Caibidil 15).

Áiseanna don oideachas: tá forbairt á déanamh ar ardáin foghlama agus ar chluichí idirghníomhacha leis na glórtha sintéiseacha. San ardán iCALL (ardán ‘cliste’ don fhoghlaim ríomhchuidithe) An Scéalaí [17], baintear úsáid as na guthanna chun scileanna labhartha agus éisteachta a chothú ag an am chéanna le scileanna léitheoireachta agus scríbhneoireachta. D’fhoghlaimeoirí óga, tá cluiche idirghníomhach á fhorbairt, Lón don Leon [1, 15], a dhíríonn ar fheasacht fóineolaíochta agus ar bhunscileanna litearthachta. Tá leaganacha píolótacha forbartha ar chluiche oideachasúil Digichaint [18] agus ar bhota comhrá Taidhgín [19]. Tá glórtha ABAIR in úsáid in ardáin sheachtracha oideachais mar Edcite chomh maith (tá tuilleadh eolais ar áiseanna oideachasúla i gCaibidil 14).

Sna na háiseanna agus sna cluichí idirghníomhacha seo dírítear ár n-aird ar thosaíochtaí forbartha na sintéise don todhchaí. Mar shampla, beidh slua carachtar agus glórtha páistí (sna canúintí éagsúla) de dhíth faoi choinne cluichí oideachasúla agus áiseanna rochtana. Is léir go mbeidh glórtha nuálacha sintéise de dhíth – inár féidir beocht agus mothúcháin a léiriú, ina bhfuil prosóid an ghlóir níos cosúla le prosóid an chomhrá dhaonna (féach §2.4.3 agus Caibidil 7). Tá taighde tosaigh ar siúl sa tionscadal Róbóglór i gColáiste na Tríonóide, ina ndírítear ar shamhail phrosóid an ghutha do na spriocanna seo.

8.5 A bhfuil le déanamh

Ag tógáil ar an taighde a rinneadh go dtí seo, is é an sprioc don chéad tréimhse eile deimhin a dhéanamh de go mbeidh córais ar fáil atá feiliúnach do na húsáideoirí agus do na háiseanna éagsúla atá de dhíth chun gur féidir leis an phobal iomlán tairbhe a bhaint astu. Tá sé tábhachtach freastal ar chanúintí breise, gan neamart a dhéanamh ar na cinn is mó i mbaol. Tá teicneolaíocht na hurlabhra ag athrú in aghaidh an lae, agus beidh uasghrádú agus measúnú leanúnach de dhíth agus córais nuálacha le tógáil de réir mar a thiocfas modhanna nua chun tosaigh.

Moltar leanúint den chur chuige idirdhisciplíneach mar a bhí i gceist go dtí seo. Beidh sé tábhachtach forbairt na sintéise a bheith ceangailte leis an taighde ar fhoghraíocht agus urlabhra na gcanúintí (§2.4 & Caibidil 7). Is féidir na modúil seo a athúsáid i dteicneolaíochtaí eile mar an aithint cainte agus tá siad lárnach le cinntiú go mbeidh na háiseanna don oideachas, don mhíchumas, don rochtain srl. ag feiliúint i gceart do na canúintí Gaeilge.

Beidh dlúth-chomhoibriú idir forbairt na gcóras sintéise agus forbairt na n-áiseanna de dhíth fosta. Is trí na háiseanna a scaiptear an teicneolaíocht i measc an phobail. Lena chois sin, is ar bhunús na n-áiseanna is fearr is féidir measúnú a dhéanamh ar na glórtha sintéiseacha (féach mar shampla [20]). Is ar bhunús riachtanais an phobail d’áiseanna ar leith (mar shampla, gléasanna dóibh siúd faoi mhíchumas radhairc agus urlabhra) is fearr a roghnaítear tosaíochtaí na teicneolaíochta. (Mar shampla, ba ar mhaithe leis an ardluas cainte a bhí riachtanach don léitheoir scáileáin do na daill a dhírigh ABAIR ar dtús ar an inneall labhartha SPSS). Amach anseo, nuair a dhéanfar nasc idir an tsintéis agus an aithint cainte (Caibidil 9), beidh béim ar chórais ‘chliste’

idirghníomhacha (Caibidil 10) d'úsáideoirí faoi leith. Beidh tionchar aige sin ar spriocanna na forbartha, fiú sa ghearrthréimhse.

Is léir go mbeidh tacaíocht bogearraí de dhíth lena chinntiú go bhfuil sintéis chainte na Gaeilge leabaithe i ngléasanna agus in aipeanna agus go ndéanfar uasdátú leanúnach orthu de réir mar a thiocfas forbairt ar an teicneolaíocht féin agus athrú ar na córais oibriúcháin.

8.6 Moltaí

Don todhchaí, moltar:

1. 1. Forleathnú na sintéise: canúintí breise, glórtha páistí, sintéis dhátheangach: le freastal ar réimse leathan úsáideoirí agus áiseanna sa todhchaí, beidh gá le:
   Canúintí breise: ag freastal ar na canúintí uile. Tá práinn faoi leith leis na canúintí atá i mbaol.
   Acmhainní teangeolaíochta agus corpais do gach canúint, le hábhair atá oiriúnach don chanúint, don chainteoir, a chlúdaíonn na fuaimeanna i ngach comhthéacs srl. Tá gá le hollchorpais labhartha le líon mór cainteoirí chun Taisce Glórtha as gach canúint a sholáthar. Mar a thagann forbairt ar na háiseanna a mbeifear ag díriú orthu, beidh corpais breise de dhíth.
   Glórtha páistí, aoisghrúpaí éagsúla agus rogha leathan glórtha: tá glórtha páistí ríthábhachtach d'áiseanna foghlama agus rochtana. Idir óg agus aosta, beidh rogha fhorleathan glórtha de dhíth, ag a bhfuil cáilíochtaí éagsúla gutha.
   Sintéis dhátheangach: atá in ann déileáil le códmhalartú (focail agus frásaí Béarla i dtéacsanna Gaeilge) agus le suíomh idirlín atá dátheangach.

Do na gnéithe seo, féach pointe 3 thíos.

2. Forbairt leanúnach ar innill labhartha agus ar mhodhanna sintéiseacha: beidh forbairt de dhíth chun a chinntiú go bhfuil na modhanna sintéise is nuálaí ar fáil. Beidh sé tábhachtach fosta aschur an taighde i moltaí 1 agus 3 a chur i bhfeidhm go héifeachtach sna córais úra seo.

3. Glórtha le beocht agus mothúcháin agus claochlú gutha: sa todhchaí beidh sintéis le glórtha solúbtha de dhíth (Caibidil 7) inar féidir cáilíochtaí prosóide an ghlóir a smachtú, agus atá feiliúnach do chórais idirghníomhacha ina mbíonn caidreamh againn le róbait, botaí comhrá agus carachtair fhíorúla i gcluichí (Caibidil 10, 13, 14, 15). Beidh gá le:
   Glórtha le beocht agus mothúcháin: cé go bhfuil córais sintéise an lae inniu intuigthe agus nádúrtha, ní féidir an glór a smachtú chun meon agus mothúcháin a chur in iúl, rud a bheidh de dhíth i gcórais idirghníomhacha chomhrá (§2.4.3 & §7.3). Tá an-tábhacht leis seo dóibh siúd le míchumas a bhaineann úsáid as an tsintéis mar a nglór féin.
   Claochlú gutha: nuair is féidir tréithe an ghutha a smachtú, is féidir fosta sraith de ghlórtha atá iomlán difriúil lena chéile a ghiniúint as aon ghlór sintéise amháin. D'fhágfadh sé seo go bhféadfadh seo líon mór carachtar a bheith ann, mar a bhíonn de dhíth do chluichí srl. D'fhágfadh sé fosta go bhféadfaí mionathruithe a chur ar cháilíochtaí an ghlóir sintéise do dhaoine faoi mhíchumas – chun glór a chruthú atá ag teacht lena bhféiniúlacht phearsanta. Nuair a bheidh 'Taisce Glórtha' le líon mór cainteoirí as na canúintí éagsúla ar fáil, beidh daoine atá faoi mhíchumas in ann glór a roghnú ar a mbunófaí an guth sintéiseach don áis atá in úsáid acu.

4. Comhpháirtíocht le forbairt na haitheanta cainte: tá taighde ar réimsí na sintéise agus na haitheanta cainte ag druidim le chéile agus tá méadú ar na scileanna innealtóireachta agus ar na hacmhainní atá i gcoiteann acu (féach Caibidil 9). Cé go bhfuil gnéithe atá an-difriúil i gceist, is léir go rachaidh comhpháirtíocht eatarthu, m.sh. i ndearadh agus i mbailiú corpas, i dtógáil acmhainní, srl., chun tairbhe don dá réimse.

5. Áiseanna píolótacha a fhorbairt ina gcórais sheasmhacha: is leaganacha píolótacha iad cuid de na háiseanna atá luaite thuas in §8.4.2, m.sh. cluichí idirghníomhacha do theagasc na Gaeilge. Tá sé tábhachtach a leithéid seo a fhorbairt ina gcórais chríochnúla sheasmhacha, tástáil a dhéanamh orthu agus iad a chur ar fáil do scoileanna, do ghrúpaí faoi leith agus don phobal i gcoitinne. Beidh comhpháirtíocht leis an phobal, le lucht proifisiúnta i réimsí an oideachais, an mhíchumais, srl. de dhíth, agus comhoibriú le heagrais Stáit chun a chinntiú go mbeidh na háiseanna ar fáil go forleathan agus tacaíocht ag lucht a n-úsáide.

6. Tástáil: beidh tástáil agus measúnú leanúnach de dhíth ar na córais agus ar áiseanna sintéisbhunaithe.

7. Pacáistiú, uasdátú, agus cothabháil ar chomhéadain do na glórtha sintéise: má tá an Ghaeilge labhartha le bheith ar fáil go forleathan i dteicneolaíochtaí éagsúla, beidh comhéadain de dhíth chun na glórtha a cheangal le gléasanna seachtracha éagsúla agus beidh pacáistiú ar áiseanna de dhíth do chórais oibriúcháin éagsúla. Le luas forbartha na teicneolaíochta agus toisc go mbíonn na córais oibriúcháin ag athrú de shíor, beidh uasdátú

leanúnach de dhíth. Le forleathnú na n-áiseanna, beidh tacaíocht theicniúil de dhíth ar lucht a n-úsáide.

Achoimre ar na moltaí

1. Leathnú ar na glórtha sintéise chun freastal ar na canúintí uile, le rogha leathan glórtha, ina bhfuil guthanna páistí agus guthanna dhátheangacha san áireamh. Beidh gá le bailiúchán ar chorpais fhorleathana, mar aon le forbairt ar acmhainní foghraíochta agus urlabhra do na canúintí (§2.4).
2. Forbairt leanúnach ar na hinnill labhartha agus ar mhodhanna nuálacha, ar bhealaí a chuireann na haschuir as 1 & 3 i bhfeidhm.
3. Díriú ar mhodhanna le glórtha solúbtha, le beocht agus mothúcháin a ghiniúint, trí chomhoibriú le taighde ar an ghlór (Caibidil 7 & §2.4.3).
4. Comhpháirtíocht le forbairt na haitheanta cainte (Caibidil 9), le tairbhe a bhaint as réimsí atá i gcoiteann acu.
5. Forbairt na sintéise a cheangal le forbairt ar áiseanna ina mbaintear úsáid as na glórtha. Moltar na háiseanna píolótacha a fhorbairt ina n-áiseanna iomlána.
6. Tástáil agus measúnú leanúnach ar ghlórtha sintéiseacha agus ar áiseanna atá bainteach leo sa dóigh is go mbeidh siad seasmhach go leor lena gcur ar fáil don phobal.
7. Uasdátú leanúnach ar chomhéadain na n-áiseanna seachtracha agus ar phacáistiú na nglórtha de réir mar a thagann athruithe ar na córais oibriúcháin. Beidh tacaíocht leanúnach de dhíth ar lucht úsáide na n-áiseanna.

Tagairtí

[1] Ní Chasaide, A., Ní Chiaráin, N., Berthelsen, H., Wendler, C., Murphy, A., Barnes, E. and Gobl, C. (2019). Leveraging phonetic and speech research for Irish language revitalisation and maintenance. Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, pp. 994-998.

[2] Ní Chasaide, A., Ní Chiaráin, N., Berthelsen, H., Wendler, C., & Murphy, A. (2015). Speech technology as documentation for endangered language preservation: The case of Irish. Paper presented at the The XVIIIth International Congress of Phonetic Sciences, Glasgow, Scotland.

[3] Taylor, P. (2009). Text to Speech Synthesis. Cambridge University Press

[4] Kominek, J., & Black, A. W. (2004). The CMU Arctic Speech Databases, Paper presented at the SSW5-2004, Pittsburg, USA.

[5] Hunt, A., and Black, A.W., (1996). Unit selection in a concatenative speech synthesis system using a large speech database, in Proceedings of ICASSP, Volume 1, pages 373-376.

[6] Zen, H., Tokuda, K. and Black, A. (2009): Statistical Parametric Speech Synthesis. Speech Communication 51:1, pp. 1-23.

[7] Tokuda, K., Tomokie, T. and Yamagashi, J. (2013) Speech Synthesis Based on Hidden Markov Models, in Proceedings of the IEEE 101:5, pp. 1234-1252

[8] Watts, O., Henter, G. E., Merritt, T., Wu, Z. and King, S. (2016) From HMMs to DNNs: Where do the improvements come from? in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China)

[9] Zen, H., Senior, A., & Schuster, M. (2013). Statistical parametric speech synthesis using deep neural networks, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[10] Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerrv-Ryan, R., Saurous, R. A., Agiomvrgiannakis, Y., & Wu, Y. (2018). Natural TTS Synthesis by Conditioning Wavenet on MEL Spectrogram Predictions. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2018-April, 4779–4783.

[11] Ren, Y., Ruan, Y., Tan, X., Qin, T., Zhao, S., Zhao, Z., & Liu, T.-Y. (2019). FastSpeech: Fast, Robust and Controllable Text to Speech. Advances in Neural Information Processing Systems, 32(NeurIPS).

[12] Łańcucki, A. (2020). FastPitch: Parallel Text-to-speech with Pitch Prediction. ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021-June, 6588–6592.

[13] Ren, Y., Hu, C., Tan, X., Qin, T., Zhao, S., Zhao, Z., & Liu, T.-Y. (2020). FastSpeech 2: Fast and High-Quality End-to-End Text to Speech. 1–15. http://arxiv.org/abs/2006.04558

[14] McGuirk, R. (2015). Exploration of the Use of Irish Language Synthesis with a Screen-Reader in the Teaching of Irish to Pupils with Vision Impairment, unpublished MPhil thesis, School of Linguistic, Speech and Communication Sciences, Trinity College, Dublin, Ireland.

[15] Ailbhe Ní Chasaide, Emily Barnes, Neasa Ní Chiaráin, Ronan McGuirk, Oisín Morrin, Muireann Nic Corcráin, and Julia Cummins. 2022. Challenges in assistive technology development for an endangered language: an Irish (Gaelic) perspective. In Ninth Workshop on Speech and Language Processing for Assistive Technologies (SLPAT-2022), pages 80–87, Dublin, Ireland. Association for Computational Linguistics.

[16] Barnes, E., Morrin, O., Ní Chasaide, A., Cummins, J., Berthelsen, H., Murphy, A., Nic Corcráin, M., O' Neill, C., Gobl, C., Ní Chiaráin, N. (2022), AAC don Ghaeilge: Prototype Development of Speech-Generating Assistive Technology for Irish, Proceedings of the CLTW 4 @ LREC2022 , pages 127–132.

[17] Ní Chiaráin, N., Ní Chasaide, A., (2019) An Scéalaí: autonomous learners harnessing speech and language technologies, SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education, Graz, Austria

[18] Ní Chiaráin, N., & Ní Chasaide, A. (2016b). The Digichaint interactive game as a virtual learning environment for Irish, in Proceedings of EUROCALL 2016, Limassol, Cyprus.

[19] Ní Chiaráin, N., & Ní Chasaide, A. (2016a). Chatbot technology with synthetic voices in the acquisition of an endangered language: Motivation, development and evaluation of a platform for Irish, in Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), Portoro, Slovenia.

[20] Ní Chiaráin, N., & Ní Chasaide, A. (2015). Evaluating synthetic speech in an Irish CALL application: influences of predisposition and of the holistic environment, in Proceedings of SLaTE 2015: 6th Workshop on Speech and Language Technologies for Education, Leipzig, Germany.

Caibidil 9
Aithint Cainte

9.1 Cad is aithint cainte ann?

Is córas é an aithint cainte (AC) ina n-aistrítear an chaint ina leagan scríofa, is é sin ó chainn go téacs. Téann sé lámh ar lámh leis an tsintéis chainte, a aistríonn téacs go caint. Le AC, is féidir téacs a chumadh ó bhéal, agus baintear úsáid as le leabhair a scríobh, nótaí a bhreacadh agus próiseas na scríbhneoireachta a bhrostú. Tá sé in úsáid go forleathan acu siúd a bhfuil disléicse orthu agus tugann sé tacaíocht cumarsáide do mhórán faoi mhíchumas, agus rochtain dóibh ar an tsaol shóisialta, ar chúrsaí oideachais agus gairme. Baintear úsáid as le gléasanna leictreonacha a reáchtáil ó bhéal. In ionaid glaonna, is minic nach duine, ach ríomhaire, a bhíonn ag éisteacht leat.

Baintear úsáid as córais aitheanta cainte agus as córais sintéise cainte, ní hamháin mar chórais iontu féin, ach fosta i gcórais choimpléascacha, mar 'chúntóir pearsanta' fíorúil, ina labhraítear le meaisín le faisnéis a chuardach agus le haiseolas labhartha a fháil (Caibidil 10). Mar chóras ann féin, nó nasctha le teicneolaíochtaí eile, tá an aithint cainte ag éirí níos forleithne in aghaidh an lae.

9.2 Cén tábhacht atá leis agus cé dó é?

Tá tábhacht faoi leith ag baint le haithint cainte don Ghaeilge: cosúil leis an tsintéis, agus go háirithe nuair a chuirtear le chéile iad, ceadaítear áiseanna a thacaíonn le forleathnú na teanga labhartha inár saol laethúil, inár gcuid oibre, inár gcaitheamh aimsire, i bhfoghlaim na teanga. Baintear úsáid as an aithint cainte in áiseanna le haghaidh:

Deachtú ríomhchuidithe: córas é seo lena gcumtar téacs ó bhéal, gan gá le méarchlár. Tá a leithéid an-áisiúil don phobal go ginearálta, dóibh siúd atá ag obair trí mheán na Gaeilge, scríbhneoirí, rúnaithe, státseirbhísigh, srl. Tá sé úsáideach d'fhoghlaimeoirí ag gach leibhéal chun aistí agus obair bhaile a ullmhú – go mór mhór nuair a cheanglaítear le gramadóirí agus le litreoirí é. Is áis thábhachtach é dóibh siúd nach bhfuil scríobh na Gaeilge ar a gcumas (go fóill) acu – agus ríthábhachtach dóibh siúd a bhfuil disléicse nó deacrachtaí luadracha nó radhairc ag gabháil dóibh.

Nótaí a bhreacadh, nuair atá lámha agus súile gafa: is uirlisí cosúla iad seo, ina mbreactar nótaí scríofa nuair atá do lámha is do shúile gnóthach – m.sh., agus tú ag tiomáint, ag láimhseáil innill, nuair atá dochtúir ag scrúdú othair, nuair atá turgnamh ar siúl ag eolaí/mac léinn.

Fotheidil uathoibríocha: le haithint cainte is féidir an focal labhartha a thaispeáint mar théacs ar scáileán, nuair atá tú ag amharc ar/éisteacht le clár ar an teilifís, ar an raidió nó ar an ríomhaire. Is féidir, mar an gcéanna, fotheidil a chur ar fáil d'aon fhoinse chainte (m.sh. léacht poiblí). Tá sé seo úsáideach don phobal, i gcúrsaí oideachais agus ríthábhachtach dóibh siúd faoi mhíchumas éisteachta.

Reáchtáil ó bhéal ar ghléasanna leictreonacha: is féidir teilifís, raidió, ríomhaire, fón póca, córas fuaime a smachtú ó bhéal. Is féidir ríomhchláir nó aipeanna a chuardach, córais teasa agus gléasanna leictreonach baile a reáchtáil srl., gan gá le comhéadain chasta, le cianrialtáin ná le méarchláir.

Cúntóir fuaimnithe ríomhchuidithe (CFR): sna mórtheangacha, tá córais aitheanta cainte á n-úsáid chun fuaimniú na teanga a theagasc. Cuirtear fuaimniú an fhoghlaimeora i gcomparáid leis an fhuaimniú dhúchasach agus tugtar aiseolas don fhoghlaimeoir chun é a cheartú.

Monatóireacht sláinte: cé nach gnáthchuid de chóras aitheanta cainte atá i gceist anseo, tá áiseanna á bhforbairt ina n-aithnítear tréithe cainte, m.sh. tréithe prosóide a bhaineann le léiriú meoin agus mothúcháin (Caibidil 2.4 & 7). Ón eolas seo is féidir monatóireacht a dhéanamh ar shláinte, agus go háirithe ar mheabhairshláinte, an chainteora, agus dáthoradh sin is féidir teiripe thráthúil a sholáthar dóibh siúd atá i mbaol.

Nuair a nasctar córas aitheanta cainte leis an tsintéis chainte agus le teicneolaíochtaí eile, cumasaítear mórán áiseanna eile, mar shampla:

Córais chomhrá ríomhchuidithe: amach anseo beidh muid ag comhrá le gléasanna éagsúla, le róbait agus le carachtair fhíorúla – comhthéacs sároiriúnach don fhoghlaim teanga (féach Caibidlí 10, 14). Tá féidearthachtaí faoi leith acu seo do dhaoine faoi mhíchumas, don chumarsáid agus chun a dtimpeallacht a smachtú (Caibidil 15). Tá róbait á bhforbairt chun cuidiú le seandaoine agus le daoine faoi mhíchumas fanacht neamhspleách ina mbaile féin.

Aistriú uathoibríoch ó chaint go caint: cé gur ó théacs go téacs a oibríonn córais aistriúcháin uathoibríocha

an lae inniu, amach anseo beidh na córais sin in ann teachtaireacht labhartha i dteanga amháin a aistriú agus a labhairt amach i dteanga eile.

9.3 Cén dóigh a n-oibríonn sé?

Tá dhá shaghas córais AC ann: (i) córas modúlach, ina bhfuil samhail acústaice, samhail teanga agus foclóir fuaimnithe, mar a léirítear i bhFíor 9.1 agus a phléitear sa chéad mhír eile, agus (ii) córas Ceann go Ceann (CgC), nach mbraitheann ar na hacmhainní sin, agus nach bhfuil ann ach aon phróiseas amháin. Cé go bhfaightear torthaí sármhaithe le córais CgC ina lán teangacha, tá siad ag brath ar chorpais ollmhóra chainte chun an córas a thraenáil agus ar ríomhairí ollchumhachtacha. I gcás mionteangacha nach bhfuil na hacmhainní sin acu, is fearr a oibríonn córais mhodúlacha, agus is iad is minice a bhíonn in úsáid.

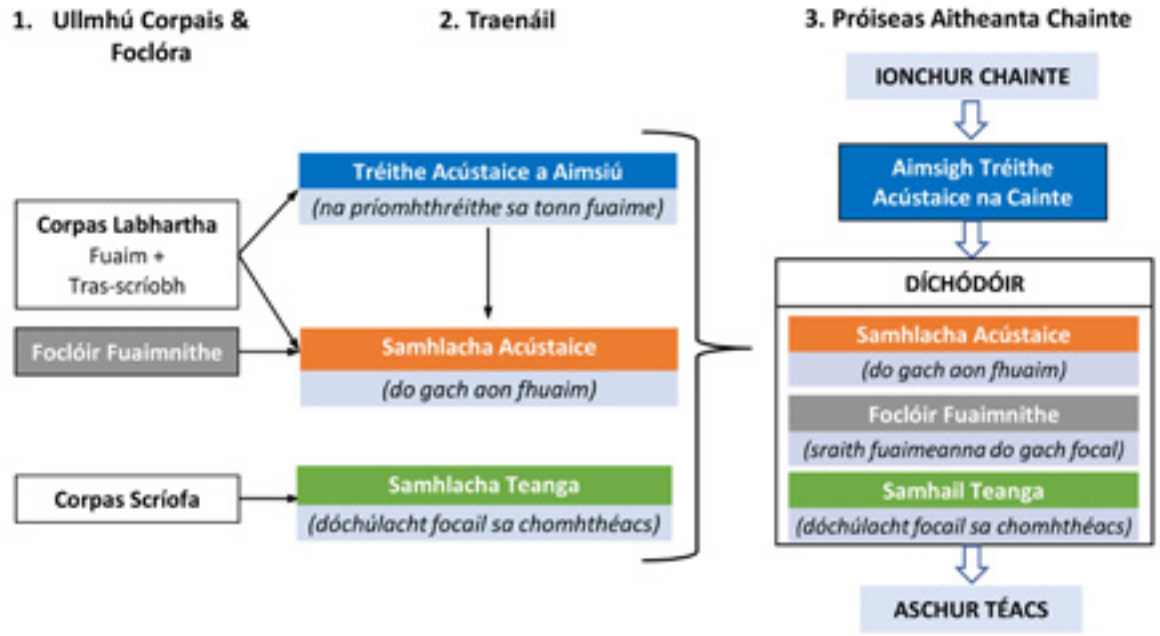
Tá cur síos níos teicniúla ar an aithint cainte in [1-3].

9.3.1 Córas Aitheanta Cainte Modúlach

Léirítear i bhFíor 9.1 na trí chéim a bhíonn i gceist agus córas aitheanta cainte modúlach á thógáil.

1 Ullmhú corpais agus foclóirí fuaimnithe
Corpais Labhartha: don aithint cainte tá gá le corpais labhartha mhóra, lena bhfuil an téacs san áireamh. Bíonn neart samplaí de dhíth d'fhuaimeanna uile na teanga, i ngach comhthéacs ina bhfaightear iad [4]. Bíonn taifead ar líon mór cainteoirí de dhíth: dá mhéad líon na gcainteoirí, is amhlaidh is fearr a bheas an córas aitheanta cainte.
Foclóir Fuaimnithe: cuireann sé seo tras-scríobh foghraíochta ar fáil do gach focal, agus do gach leagan éagsúil d'fhocal. Is fada foclóirí fuaimnithe ann don Bhéarla agus do na mórtheangacha. Don Ghaeilge, níl foclóir fuaimnithe ann atá oiriúnach don aithint cainte agus is gá iad a chruthú as an nua. Toisc nach bhfuil tús áite ag aon chanúint amháin mar chanúint chaighdeánach labhartha, tá foclóirí fuaimnithe de dhíth do na canúintí éagsúla.
Corpais Scríofa: baintear úsáid fosta as corpais mhóra d'ábhar scríofa.

2 Traenáil bunaithe ar na corpais/foclóirí fuaimnithe
Baintear úsáid as modhanna meaisínfhoghlama staitistiúla (féach Caibidil 1.7) chun páirteanna éagsúla



Fíor 9.1: Na céimeanna a bhaineann le córas aitheanta cainte modúlach a thógáil

den chóras a thógáil ar bhunús na gcorpas agus na bhfoclóirí fuaimnithe. Leis an taifead fuaime agus an tras-scríobh fuaime air, cruthaítear Samhail Acústaice de na príomhthréithe atá ag gach fuaim (gutaí agus consain). Toisc go bhfuil an-tionchar ag fuaim amháin ar na cinn roimhe agus ina dhaidh (Caibidil 2.4), tógtar samhail acústaice do gach comhthéacs ina bhfaightear an fhuaim. Baintear úsáid as an mórchorpas téacs (chomh maith leis na leaganacha scríofa de na corpais labhartha) chun Samhail Teanga a thógáil – go minic i bhfoirm Samhail N-Gram. Is éard atá i gceist leis seo tábla ina bhfuil minicíocht gach focail, ag brath ar na focail a tháinig roimhe (de ghnáth, 2 nó 3 fhocal). Mar sin, le linn dhíchódú na cainte, ceadaíonn an tsamhail teanga roghnú idir na focail éagsúla a d'fhéadfadh freagairt don tsraith fuaime atá aimsithe (féach thíos). Le blianta beaga anuas tá samhlacha teanga eile ag teacht chun tosaigh, ina mbaintear úsáid as líonraí néaracha domhainfhoghlama, agus inar féidir na focail roimhe agus ina dhiaidh a chur san áireamh.

3 Díchódú: aithint na cainte

Nuair a labhraíonn duine leis an chóras aitheanta cainte, sa chéad áit aimsítear tréithe acústaice na cainte agus seoltar chuig an díchódóir iad. Don díchódú, meastar dóchúlacht na sraitheanna fuaime agus na bhfocal a d'fhéadfadh freagairt do na tréithe acústaice atá aimsithe: dó seo baintear úsáid as na samhlacha acústaice (faoi choinne fuaimeanna i gcomhthéacsanna éagsúla) agus an tsamhail teanga (do na focail is minice a mbeadh súil leo sa chomhthéacs). Tá an foclóir fuaimnithe mar nasc idir sraitheanna fuaime agus na focail. Roghnaítear agus cuirtear in ord na focail agus na sraitheanna fuaime is fearr a fhreagraíonn don ionchur cainte. Mar aschur ón díchódóir, pioctar an ceann is cosúla: i gcórais áirithe tugtar níos mó ná rogha amháin don úsáideoir.

9.3.2 Córas Ceann go Ceann (CgC)

Leis an dul chun cinn ar líonraí néaracha domhainfhoghlama (DNN) [3] agus le méadú ar chumhacht na ríomhairí le blianta beaga anuas, tá feabhas tapa ag teacht ar chruinneas na gcóras aitheanta cainte [5-7]. Nuair a bheidh corpais labhartha ollmhóra ann do na canúintí, beidh féidearthachtaí córais nuálacha a chur ar fáil don Ghaeilge mar atá ag teacht chun tosaigh i dteangacha ina bhfuil a leithéid ar fáil. Ina leithéid seo de chóras, déantar aithint ar gach gné chainte in aon phróiseas amháin, ina bhfoghlaimíonn na líonraí néaracha an nasc idir an tonn fuaime agus an téacs, gan bheith ag brath ar na hacmhainní thuasluaite [8-10]. Is minic a bhaintear úsáid go hindíreach as acmhainní mar fhoclóirí fuaimnithe agus samhlacha teanga mar sin féin. Is fiú a lua fosta go bhfuil na hacmhainní teanga tábhachtach do mhórán áiseanna ina mbainfear úsáid as an aithint cainte.

I gcás na Gaeilge, tá dúshláin faoi leith ag baint le háiseanna aitheanta cainte a sholáthar a fhreagraíonn go hiomlán do riachtanais fhorleathana phobal na Gaeilge. Mar a phléitear i gCaibidil 1, ní hionann do mhionteanga mar an Ghaeilge agus mórtheangacha mar an Béarla, ina bhfuil tosaíochtaí na teicneolaíochta ag brath ar mhórghnólachtaí agus ar bhrabús a gcuid earraí. Beidh sé tábhachtach go mbeidh tuiscint ar phobal agus ar chomhthéacs sóisialta na teanga ag stiúradh thosaíochtaí na forbartha, agus go mbeidh an bhéim ar na háiseanna is riachtanaí don phobal agus do chaomhnú na teanga. Pléitear é seo i mír §9.5.

9.4 A bhfuil déanta i leith na Gaeilge go dtí seo

Tá forbairt ar an aithint cainte ar siúl ag foireann ABAIR i gColáiste na Tríonóide agus tá córas modúlach béite, ÉIST, ar fáil ar an suíomh gréasáin . Tá córas curtha ar fáil fosta le déanaí ag Microsoft, mar chuid de phacáiste ilteangach. I dtógáil an chórais ÉIST, baineadh úsáid sa chéad áit as na corpais le tras-scríobh foghraíochta, a bhí ullmhaithe cheana féin don tsintéis (Caibidil 8), chomh maith le taifeadtaí ar chanúintí éagsúla ón ardán sluafhoinsithe MíleGlór, agus le corpais labhartha eile a bhí ar fáil. Baineadh úsáid fosta as acmhainní sintéiseacha eile (Caibidil 8), m.sh. foclóirí fuaimnithe agus na rialacha litir-go-fuaim do na trí mhórchanúint. Chun freastal ar éagsúlacht na gcanúintí, tógadh foclóir fuaimnithe traschanúnach ag baint úsáid as coincheap le Ó Murchú [11] faoi 'chroílár comónta' i gcórais fuaime na gcanúintí Gaeilge. Tógadh na samhlacha teanga i gcomhpháirt le Fiontar in Ollscoil Chathair Bhaile Átha Cliath, ag baint úsáid as a gcuid corpas scríofa (féach Caibidil 3). Tá cur síos ar an chóras ÉIST in [12].

Tá córais CgC á dtógáil ag an am chéanna, ach níl an cruinneas iontu (go fóill) is atá sa chóras mhodúlach. Braitheann na córais seo ar chorpais labhartha ollmhóra agus táthar ag díriú faoi láthair ar ullmhú agus ar thaifeadadh corpais, go háirithe sna ceantair Ghaeltachta, sa dóigh is gur féidir córais nuálacha don aithint agus don tsintéis chainte a sholáthar do na canúintí uile. Tá ábhair

labhartha á mbailiú fosta atá taifeadta cheana féin. Sa chás nach bhfuil téacs ar fáil leis na comhaid fuaime, tá úsáid á baint as samhlacha leathstiúrtha don trascríobh. Tá córais fhéinmhaoirsithe réamhthraenáilte in úsáid fosta. Le traenáil agus le tiúnáil leanúnach don Ghaeilge, tá feabhas ag teacht ar an aschur de réir a chéile.

Tá áis le haithint cainte á tógáil i gcomhpháirt le TG4, ina n-aithneofar eochairfhocail chun gur féidir cláir theilifíse a aimsiú agus seinnteoir TG4 a rialú ó bhéal. Tá trialacha ar siúl chun an aithint cainte a leabú in áiseanna ina bhfuil an tsintéis chainte ar fáil cheana féin (Caibidil 8) agus tá béim faoi leith ar áiseanna oideachasúla agus rochtana.

9.5 A bhfuil le déanamh

I gcás na Gaeilge, is é an dúshlán is mó ná córais aitheanta cainte a sholáthar a fhreastalóidh i gceart ar phobail dhifriúla na Gaeilge, .i. córais atá in ann déileáil le héagsúlacht na gcainteoirí, agus leis na comhthéacsanna ina mbainfear úsáid astu. Sna mórtheangacha mar an Béarla, is ar an chanúint chaighdeánach a bunaíodh na chéad chórais, agus leathnaíodh amach ina dhiaidh sin iad. I gcás na Gaeilge, is beag an mhaith córas nach n-aithneodh cainteoirí dúchais as gach canúint. Caithfidh na córais chéanna caint na nua-Ghaeilgeoirí a aithint fosta. Agus, mar gheall ar thábhacht na teicneolaíochta d'áiseanna oideachais, beidh córais de dhíth a aithníonn caint ó fhoghlaimeoirí chomh maith, ag leibhéil dhifriúla chumais – rud atá dúshlánach. Is léir ó thaighde a rinneadh ar theangacha eile gur measa go mór a aithnítear an chaint ag foghlaimeoirí i gcomparáid le cainteoirí dúchais [13].

Mar gheall ar thábhacht na teicneolaíochta do theagasc na teanga, beidh sé ríthábhachtach go mbeidh córais ann atá ábalta glórtha páistí ag aoiseanna difriúla a aithint, chomh maith le daoine fásta, idir fhir agus mhná. Mar a míníodh i gCaibidil 2.4.2, braitheann tréithe acústaice na bhfuaimeanna ar mhéid an bhéil agus na mball urlabhra, rud a athraíonn faoi mar a fhásann páiste ó bhliain go bliain. Fiú sna teangacha ina bhfuil na córais is nuálaí tá dúshláin ar leith ag baint le glórtha páistí.

Tá tionchar ag tréithe teangeolaíochta na teanga ar na bunriachtanais don aithint chainte [14], m.sh. ar mhéid is ar dhearadh na gcorpas, is ar na foclóirí fuaimnithe atá de dhíth le córas bunúsach a thógáil. Mar sin, tá impleachtaí ag an réimse fairsing d'fhuaimeanna sa Ghaeilge (Caibidil 2.4.2) agus líon na leagan infhillte d'fhocail (foirmeacha éagsúla focail de réir na gramadaí, Caibidil 2.5.1), go háirithe nuair a chuirtear na canúintí éagsúla san áireamh (féach an plé ar na gnéithe seo in [12]). I gcás áiseanna ar leith, beidh dúshláin a baint leis an chódmhalartú atá forleathan sa Ghaeilge (i.e., focail Bhéarla sa chaint [15, 16]).

Tá dúshláin choiteanna eile i gceist. Faoi choinne córas seasmhach is féidir a úsáid i suíomhanna éagsúla, caithfear an chaint a aithint fiú nuair a bhíonn callán sa chúlra (m.sh. i seomra ranga) nó nuair a athraíonn stíl na cainte – m.sh. ó chaint cúramach go comhrá neamhfhoirmiúil gasta.

Sa todhchaí, beidh béim ar an aithint cainte mar pháirt de chóras comhrá, ina mbíonn caidreamh againn le meaisín, m.sh. i bhfoirm cúntóir pearsanta nó carachtair fhíorúla i gcluiche. Tá dúshláin ar leith ag baint lena leithéid de chórais, sa mhéid is go ndéantar aithris ar an chomhrá dhaonna. Mar a míníodh i gCaibidil 2.4.3 agus i gCaibidil 7, braitheann brí na cainte ní hamháin ar na focail (rud a aithníonn an córas aitheanta cainte) ach ar an dóigh a ndeirtear iad chun meon agus mothúcháin an chainteora a nochtadh (trí athruithe prosóide an ghutha, rud nach n-aithníonn an córas aitheanta). Nuair a bhíonn dea-chaidreamh idir beirt, éiríonn a bpatrúin prosóide níos cosúla lena chéile: is slat tomhais é seo ar éifeacht na cumarsáide eatarthu [13]. Tá taighde na hurlabhra ag díriú níos mó agus níos mó ar na gnéithe tábhachtacha seo cainte nach scríobhtar, agus tá fás ag teacht ar réimse ríomheolaíochta na mothúchán agus ar áiseanna nuálacha mar an mhonatóireacht sláinte atá luaite thuas. Sa todhchaí, caithfidh na córais aitheanta cainte a bheith in ann níos mó ná na focail a thagann as béal duine a aithint.

Cé gur céim ar chéim a thabharfar aghaidh ar na dúshláin seo uile, ó thaobh na pleanála de is fiú meabhrú ó thús ar an mhórphictiúir, agus a chinntiú go bhfuil na spriocanna gearrthéarmacha do na chéad chórais ag soláthar na dúshraithe don fhorbairt fhadtéarmach. Mar an gcéanna le teicneolaíochtaí eile, moltar forbairt chomhthreomhar a dhéanamh ar bhunteicneolaíocht na haitheanta cainte agus ar áiseanna ina bhfuil sí leabaithe, chun go mbeidh aschur an taighde á chur ar fáil go leanúnach don phobal.

9.6 Moltaí

I gcomhthéacs na riachtanas gearrthéarmach agus fadtéarmach a bhaineann le haithint cainte na Gaeilge, moltar:

1. Forbairt na haitheanta cainte a nascadh le forbairt na sintéise. Tá corpais labhartha lárnach don dá réimse, agus pléitear an ghné seo in 2 thíos. Lena chois sin, tá na réimsí taighde ag druidim ar a chéile,

agus na féidearthachtaí ag méadú do chomhthiomsú saineolais, scileanna agus acmhainní a tacaíonn leis an dá réimse. Is iomaí dúshlán i gcoiteann acu, agus is fiú straitéis choiteann forbartha a chur i bhfeidhm, a chinntíonn go bhfuil teicneolaíochtaí na hurlabhra ag freastal mar is cóir ar na pobail éagsúla a bheidh ag brath orthu (cainteoirí dúchais gona gcanúintí éagsúla, nua-Ghaeilgeoirí, foghlaimeoirí, páistí, srl.), agus ag díriú de réir a chéile ar na háiseanna is mó a mbeidh gá leo (m.sh. córais chomhrá le prosóid nádúrtha, stíleanna éagsúla cainte, códmhalartú, srl.).

2. Corpais labhartha ar scála ollmhór a fhorbairt: ar deireadh thiar thall, braithfidh caighdeán agus éifeacht na gcóras ar chaighdeán, líon agus meascán na gcainteoirí sna corpais ar a bhfuil siad bunaithe. Cé nach ionann do chorpais shintéiseacha agus aitheanta cainte (maidir le méid na gcorpas, líon na gcainteoirí, an trealamh/seomra taifeadta), mar atá ráite, tá mórán dúshlán i gcoiteann acu, agus moltar go mbeadh taighdeoirí sa dá réimse ag obair as lámha a chéile. Do bhailiú corpais, moltar straitéis fhadtéarmach agus comhpháirtíocht le pobail na Gaeilge, ina dtugtar aird ó thús ar éagsúlacht na gcanúintí, na n-aoisghrúpaí, na stíleanna cainte, an timpeallacht taifeadta, srl. atá de dhíth. Maidir le dearadh na gcorpas, beidh gá le clúdach ar na fuaimeanna uile i ngach comhthéacs, do gach canúint; beidh gá le hábhar don taifeadadh atá in oiriúint do na canúintí, do na haoisghrúpaí éagsúla agus do na stíleanna cainte a oireann d'áiseanna ar leith – m.sh. ábhar a léitear go cúramach nó comhrá neamhfhoirmiúil. Is féidir úsáid a bhaint fosta as an méid mór ábhar labhartha atá ar fáil. Nuair nach bhfuil an leagan scríofa ar fáil leis na taifeadtaí, is féidir úsáid a bhaint as samhlacha féinmhaoirsithe nó leathstiúrtha i dtraenáil an chórais aitheanta cainte.

3. Tógáil agus tástáil leanúnach ar chórais aitheanta cainte de réir mar a thagann méadú ar na corpais. Tá gá le forleathnú ar fhoclóirí fuaimnithe, rialacha litir-go-fuaim do gach uile canúint, ar shamhlacha acústaice (bunaithe ar chorpais labhartha) agus ar shamhlacha teanga (bunaithe ar chorpais téacs) chun na córais mhodúlacha a fheabhsú. Faoi mar a fhásann na corpais atá inúsáidte, moltar córais Ceann go Ceann a thógáil ag baint úsáid as samhlacha réamhthraenáilte.

4. Áiseanna – inar féidir córais shrianta a úsáid: is é an mhórsprioc ná córas foirfe a chur ar fáil lena n-aithnítear cainteoirí éagsúla (páistí, mná, fir, cainteoirí dúchais gona gcanúintí éagsúla, foghlaimeoirí, srl.), le caint nádúrtha neamhfhoirmiúil, fiú nuair atá callán sa chúlra. Mar sin féin, sa ghearrthréimhse, moltar díriú ar áiseanna ina mbaintear úsáid as córais níos srianta. De ghnáth, tógtar áiseanna do ghrúpaí agus do thascanna ar leith, rud a cheadaíonn srian a chur le réimse na haitheanta, mar a oireann don chomhthéacs agus don lucht úsáide. Tá bealaí éagsúla le barr feabhais a chur ar chórais shrianta, mar shampla:
   Foclóir agus tascanna srianta: tá a leithéid úsáideach in áiseanna oideachasúla ina bhfuil eolas roimh ré ar an fhoclóir iomlán atá de dhíth. D'fhéadfaí, mar seo, barr feabhais a chur ar áis a bheadh dírithe ar ghuthanna páistí.
   Córais shrianta d'úsáideoir(í) faoi leith: is minic in áiseanna don deachtú go ndéantar 'traenáil' ag an tús ar an chóras don té atá á úsáid. Léann sé/sí abairtí áirithe amach, agus cuireann an córas na samhlacha acústaice in oiriúint don chainteoir. D'fhéadfaí an straitéis seo a úsáid chun barr fheabhais a chur ar áis do chanúint ar leith.
   Aithint ar fhocail aonaracha: is féidir go gcuideodh sé le cruinneas na haitheanta cainte dá labhródh an cainteoir focal ar fhocal, le briseadh beag idir focail. Is féidir lena leithéid seo cabhrú i gcumadh ríomhphoist, aistí srl. agus a bheith úsáideach dóibh siúd a bhfuil disléicse nó míchumas radhairc ag gabháil dóibh.
   Eochairfhocail a aimsiú: tá áiseanna ann ina mbaintear úsáid as an aithint cainte d'eochairfhocail ar leith: le haithint focail ceanglaítear gníomhaíocht faoi leith, m.sh. gléas leictreonach a rialú. Ní gá don chóras aitheanta ach líon beag eochairfhocal a aithint, ach caithfidh sé a bheith in ann déileáil le mórán guthanna éagsúla (canúintí éagsúla, páistí, srl.). Is sampla dá leithéid an áis thuasluaite chun seinnteoir TG4 a rialú ó bhéal.

Moltar go mbeidh tosaíocht ag na háiseanna is mó a bhfuil gá leo, agus atá indéanta leis na hacmhainní atá ar fáil ag an am. Mar sin, beidh dlúthchaidreamh le pobal na Gaeilge de dhíth i roghnú na dtosaíochtaí. Agus cé go dtógfaidh sé tamall sula mbeidh aithint cainte chríochnúil ar fáil lena ndéantar freastal ar an réimse leathan cainte agus comhthéacsanna ina n-úsáidfí a leithéid, le gach céim chun tosaigh beifear in ann áiseanna suimiúla, úsáideacha agus féidearthachtaí úra don oideachas, don rochtain agus don phobal i gcoitinne a thabhairt ar an fhód.

Achoimre ar na moltaí

1. Clár taighde ina nasctar aithint agus sintéis na cainte, agus comhthiomsú ar scileanna, modhanna oibre agus acmhainní.

2. Forbairt ar chorpais ollmhóra labhartha, ina n-áirítear ó thús (i ndearadh, i mbailiú agus i bpróiseáil na gcorpas) riachtanais fhadtéarmacha na haitheanta agus na sintéise cainte, chun go mbeidh siad ag freagairt mar is cóir amach anseo do na húsáideoirí agus do na comhthéacsanna éagsúla ina mbainfear úsáid astu:
a. Cainteoirí dúchais do na canúintí éagsúla, fir, mná agus páistí d'aoiseanna difriúla
b. Nua-Ghaeilgeoirí agus foghlaimeoirí ag leibhéil éagsúla, idir óg agus aosta.
c. Stíleanna éagsúla cainte, callán sa chúlra, srl.
3. Forbairt, tástáil agus leasú leanúnach ar bhunchórais aitheanta cainte. Chomh maith le forbairt na gcorpas labhartha, beidh forbairt leanúnach de dhíth ar fhoclóirí fuaimnithe canúnacha, ar shamhlacha acústaice agus ar shamhlacha teanga.
4. Forbairt ar áiseanna srianta, do thascanna agus d'úsáideoirí faoi leith.

Tagairtí

[1] Rabiner, Juang (1993). Fundamentals of Speech Recognition. Prentice Hall.

[2] C.-C. Chiu, T. N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R. J. Weiss, K. Rao, E. Gonina, et al. (2018). State-of-the-art speech recognition with sequence-to-sequence models, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 4774–4778, 2018.

[3] Nassif, Ali Bou, et al. (2019) "Speech recognition using deep neural networks: A systematic review." IEEE access 7: 19143-19165. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8632885

[4] Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: Timit and beyond. Speech Communication, 9(4), 351–356. https://doi.org/10.1016/0167-6393(90)90010-7

[5] Jurafsy, D. and Martin, J. H. (2019) Speech and Language Processing https://web.stanford.edu/~jurafsky/slp3/ https://web.stanford.edu/~jurafsky/slp3/ed3book_dec302020.pdf

[6] Hinton, G. E., S. Osindero, and Y.-W. Teh (2006). "A fast learning algorithm for deep belief nets". Neural Computation 18.7, pp. 1527–1554.

[7] Mohamed, A.-r., G. Dahl, and G. Hinton (2009). "Deep belief networks for phone recognition". In: NIPS Workshop on Deep Learning for Speech Recognition and Related Applications. Vancouver, Canada.

[8] Yu, D. and L. Deng (2010). "Deep learning and its applications to signal and information processing". In: IEEE Signal Processing Magazine 28.1, pp. 145–154.

[9] Chorowski, I.K., Bahdanau, D. Serdyuk, D., Cho, K., Bengio, Y. (2015) Attention-based models for speech recognition, in: Advances in neural information processing systems, 577–585..

[10] W. Chan, W., Jaitly, N., Le, Q., Vinyals, O. (2016). Listen, attend and spell: A neural network for large vocabulary conversational speech recognition, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 4960–4964.

[11] Ó Murchú, M. (1969). Common core and underlying forms. Ériu, 21, 42-75.

[12] Liam Lonergan, L., Qian, M., Ní Chiaráin, N., Gobl, C., Ní Chasaide, A. (2022). Cross-dialect lexicon optimisation for an endangered language ASR system: the case of Irish, in Proceedings of Interspeech 2022, Incheon, Korea.

[13] Tomokiyo, L.M., (2001). Recognizing non-native speech: Characterizing and adapting to non-native usage in LVCSR. PhD thesis, Carnegie Mellon University, Pittsburg.

[14] Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014). Automatic speech recognition for under-resourced languages: A survey. Speech Communication, 56, 85-100.

[15] Ní Laoire, S. (2019) 'Códmhalartú, códmheascadh agus trasteangú: teangacha i dteagmháil' [Codeswitching, codemixing and translanguaging: languages in contact] in Ó hIfearnáin, Tadhg (ed.) An tSochtheangeolaíocht: Taighe agus gníomh Baile Átha Cliath: Cois Life, 49-78.

[16] Ní Laoire, S. (2016). 'Irish-English code-switching: a sociolinguistic perspective' in Hickey, Ray (ed.) Socioi dtógáil na foclóra fuaimnithe, linguistics in Ireland London: Palgrave Macmillan, 81- 106.

[17] Weise, A., Levitan, S. I., Hirschberg, J., & Levitan, R. (2019). Individual differences in acoustic-prosodic entrainment in spoken dialogue. Speech Communication, 115, 78-87.

Caibidil 10
Córais Chomhrá Ríomhbhunaithe

10.1 Cad iad?

Cuireann córais chomhrá ríomhbhunaithe (CCR) ar chumas daoine agus meaisíní idirphlé a dhéanamh trí theanga nádúrtha. In áit a bheith ag baint úsáid as ríomhchláir theicniúla nó códúcháin chasta ríomhaireachta, is féidir le daoine labhairt leis na CCR nó téacs a chur chucu ar a bhfón póca, ar a dtáibléad nó ríomhaire glúine agus freagra a fháil i dteanga nádúrtha nó mar théacs. Bogearraí casta a tharraingíonn ó theicneolaíochtaí éagsúla bunaithe ar intleacht shaorga iad na córais seo. Tá na bunchlocha don teanga nádúrtha nó do chomhéadain chainte (guthanna sintéiseacha (GS), aithint cainte agus próiseáil teanga nádúrtha) seanbhunaithe don Bhéarla agus tá siad á bhforbairt don Ghaeilge, mar atá mínithe i gCaibidlí 8, 9 agus 4 den Phlean Digiteach seo.

Bhí na luathchórais chomhrá téacsbhunaithe, ach le forbairtí i dteicneolaíocht na teanga labhartha, tá caint idir an úsáideoir agus an meaisín anois indéanta. Sampla de chúntóir digiteach pearsanta nua-aimseartha é leithéid Siri, agus baintear úsáid go forleathan as córais atá níos sofaisticiúla i réimsí mar bhaincéireacht, cúram sláinte, tráchtáil (seirbhísí miondíola/do chustaiméirí) agus oideachas. Féachtar ar chomhéadain teanga nádúrtha, agus ar chomhéadain ghutha go háirithe, mar réimsí móra forbartha sa teicneolaíocht sna blianta atá amach romhainn. Is ar an gcaint atá formhór na hidirghníomhaíochta daonna bunaithe, agus mar sin cuireann an teicneolaíocht ghutha ar chumas ríomhairí feidhmiú i gcásanna inarb í an teanga labhartha an rud is nádúrtha nó an t-aon mhodh caidrimh atá inúsáidte, mar shampla, rohtain do dhaoine faoi mhíchumas, faisnéis shiamsúil, oideachas siamsúil agus cluichí oideachasúla.

Tá CCR trasardáin agus ilmhódach – tá na córais atá in ann feidhmiú trí chaint, trí théacs nó tríd an dá cheann. Is féidir leis an gcomhéadan a bheith chomh simplí le córais teachtaireachta ar fhón póca (Facebook Messenger, WhatsApp), nó chomh casta le habhatár a bhfuil cuma ilfheidhmiúil dhaonna air, nó fiú róbat fisiciúil.

10.2 Cén tábhacht atá leo agus cé dó iad?

10.2.1 Córais reatha

Léiríonn Tábla 10.1 samplaí den úsáid a bhaintear as córais chomhrá ríomhbhunaithe faoi láthair. Is i gcúrsaí gnó is mó a úsáidtear iad, mar shampla i lárionaid glaonna. Níl tábhacht thar na bearta ag roinnt le cuid mhaith de na samplaí i gcás na Gaeilge ach tá na samplaí is mó spéis sna boscaí scáthaithe sa tábla. Tá tuilleadh plé orthu thíos.

Samplaí	Cé dó?
Carrloingseoireacht	An pobal i gcoitinne
Lárionaid glaonna	An pobal i gcoitinne: idirbhealach don phobal (gnó & institiúidí)
Deasca cabhrach	General public: interface to public (business & institutions)
(uathoibríocha)	An pobal i gcoitinne: idirbhealach don phobal (gnó & institiúidí)
Idirbhealaí sláinte	An pobal i gcoitinne: m.sh. eolas faoi shláinte phobail a thabhairt – féach Figiúr 10.1, SimSensei córas comhrá fíorúil daonna a dhéanann agallaimh a bhaineann le ciapadh siceolaíochta (mar dhúlagar buairt agus neamhord strus iarthrámach)
Comhéadain shóisialta agus lámhleabhair dialógbhunaithe	An pobal i gcoitinne: m.sh. lámhleabhair chaintbhunaithe chun scileanna nua a shealbhú, fabhtcheartú, s.rl.
Siri, Alexa, Cortana, Google Assistant, s.rl.	An pobal i gcoitinne (cuiditheoir pearsanta) / Míchumas & rochtain
Córais chliste teagaisc	An pobal i gcoitinne /Lucht foghlama
Comhpháirtithe comhrá	An pobal i gcoitinne /Lucht foghlama
Teagascóirí	Lucht foghlama
Feidhmeanna tumthacha eolais	An pobal i gcoitinne /Sainphobail amhail treoraithe músaem / cúntóirí ranga (Fíoracha 10.2, 10.3)

Tábla 10.1: Samplaí den úsáid a bhaintear as Chórais Chomhrá. I bhFíor 10.1 léirítear comhéadan reatha trínar féidir le húsáideoirí, a bhfuil neamhord siceolaíoch orthu, labhairt le speisialtóir fíorúil sa chúram sláinte.

Fíor 10.1: SimSensei & MultiSense: Duine Fíorúil agus Braistintí Ilmhódúla le haghaidh Tacaíocht Cúraim Sláinte (https://www.youtube.com/watch?v=ejczMs6b1Q4)

Córais ceisteanna & freagraí, m.sh. cuiditheoirí pearsanta fónbhunaithe: cosúil le Siri – is féidir ceist a chur, ansin cuardaíonn an córas bunachar sonraí agus tagann sé ar ais le freagra. Níl fáil air seo don Ghaeilge ach úsáidtear córais cosúil le Siri go forleathan sa Bhéarla. Tá éileamh air ag fás i gcás na Gaeilge, ó dhaoine a bhfuil lagú radhairc ag gabháil dóibh go háirithe. Tabhair faoi deara go bhfuil teacht ag córais amhail Siri ar bhunachair ollmhóra sonraí nach bhfuil ar fáil don Ghaeilge. Ina theannta sin, níl an aithint cainte don Ghaeilge ach ina tús ó thaobh forbartha de.

Córais chliste teagaisc (CCT): is féidir córais a thógáil a chuireann teagasc oiriúnaithe ar fáil d'fhoghlaimeoirí in iliomad réimsí trí úsáid a bhaint as teicneolaíocht na teanga labhartha agus próiseáil teanga nádúrtha. Cabhraíonn na córais chliste teagaisc seo le foghlaimeoirí trí cheisteanna a chur, freagraí a pharsáil, agus aiseolas le teagasc sainoiriúnaithe a thabhairt, gan cabhair ó mhúinteoirí de ghnáth. Nuair a cheaptar gréasán cuí is féidir leis na córais seo cur go mór le teagasc ranga, trí réimse ábhar a chur ar fáil go cianda agus trí dhianchleachtadh breise a chur ar fáil ar phraghas íseal nó saor in aisce. Tá córais mar seo in úsáid cheana féin agus ag éirí go maith leo i réimsí na heolaíochta fisiciúla agus na matamaitice, agus tá siad ag teacht ar an bhfód anois san eolaíocht sláinte, i sealbhú teangacha agus i réimsí eile foghlama foirmeálta.

Tugann na córais nua-aimseartha CCT aiseolas láithreach ag gach seisiún agus cabhraíonn siad le measúnú leanúnach trí phroifíl an fhoghlaimeora a choimeád cothrom le dáta. Coimeádann na córais seo taifead ar chumais na bhfoghlaimeoirí agus déanann siad réamh-mheas ar a gcuid ábaltachtaí agus ar a gcuid riachtanas. Cabhraíonn sé seo le sainchuraclam fuinniúil a chruthú do gach duine. Ríomhann an córas meastachán staitistiúil den mháistreacht atá ag an bhfoghlaimeoir ar na coincheapa agus na scileanna a ghabhann le gach ceacht. Leanann na córais seo orthu le ceisteanna a chur ar an bhfoghlaimeoir go dtí go ndéanann siad amach go bhfuil máistreacht ag an bhfoghlaimeoir ar na coincheapa. Bailíonn siad sonraí faoi gach foghlaimeoir atá á n-úsáid ag an am céanna.

Go dtí le déanaí, bhí na córais neamhphearsanta, neamhshóisialta, teagascúil agus teibí agus ní raibh sé d'aidhm acu ach eolas a roinnt. Bhí easpa tréithe pearsanta orthu den chuid is mó, rud a d'fhág nach ndearna na foghlaimeoirí caidreamh pearsanta leo, agus bac á chur ar an inspreagadh pearsanta dá réir.

Gníomhairí comhrá: Is mór idir comhrá daonna agus téacs atá á léamh amach os ard. Mar a pléadh i gCaibidlí 2, 7 agus 8, níl na córais chomhrá ríomhbhunaithe ábalta déileáil leis na gnéithe réalacha eispriseacha den chomhrá daonna ach tá an teicneolaíocht ag dul sa treo seo. Is é an comhrá nádurtha an eochair chun chun caidreamh láidir oibre a chruthú idir an córas agus an t-úsáideoir – tá sé léirithe go neartaíonn comhrá nádúrtha caidreamh gníomhach, foghlaim agus dílseacht nuair a úsáidtear é i gcúrsaí oideachais nó i gcúram sláinte. Lena leithéid seo de chórais spreagtar an t-úsáideoir chun féachaint ar an gcóras mar pháirtnéir seachas mar mheaisín go díreach. Tá tábhacht thar na bearta leis seo ó thaobh feidhmeanna oideachasúla de.

Scríobh McTear in 2002 go raibh comhrá ríomhbhunaithe mar aidhm ag lucht taighde in earnáil na teicneolaíochta cainte agus san intleacht shaorga le 30 bliain roimhe sin agus tá an aidhm sin fós romhainn amach (McTear, 2002). Go bunúsach, ligeann an gníomhaire don úsáideoir comhrá a dhéanamh le 'héisteoir' fíorúil sa ríomhaire – is é sin gníomhaire a dhéanann caint le daoine chun cabhrú leo nó chun spraoi a dhéanamh leo.

Úsáidtear an téarma 'córas comhrá ríomhbhunaithe' i réimse leathan saoráidí. Tá cuid acu seo níos cúinge ná córais oscailte chainte sa mhéid go bhfuil siad teoranta do thopaicí ar leith – fiosrúcháin eitleoireachta, mar shampla. Uaireanta ní bhíonn sa chumarsáid ach réimsí an-chúng le caint nádúrtha agus cúpla focal mar uimhreacha ó 0-9 agus 'tá' nó 'níl', ach bíonn stór mór focal ag cuid eile acu rud a cheadaíonn réimse cainte atá níos leithne. Is é an aidhm fhadtréimhseach atá ann, áfach, córas fíorchomhrá a bhunú atá in ann comhrá a choimeád faoi aon topaic faoin spéir.

Tumfhoinsí faisnéise: Tá sampla amháin léirithe in bhFíor 10.2 agus is treoraí fíorúil músaem é. Is féidir le cuairteoirí comhrá nádúrtha aonair a bheith acu leis an treoraí fíorúil, ceisteanna a chur agus freagraí a fháil. Ba mhór an

buntáiste é a bheith ábalta plé le daoine idir óg is aosta trí fhaisnéis a roinnt leo i suíomh idirghníomhach. Chuirfeadh sé seo leis an réimse suíomhanna ina dtiocfadh cainteoirí dúchais agus foghlaimeoirí ar an nGaeilge. Bheadh sé éasca go leor áis mar seo a chruthú mar tá tráchtaireacht Ghaeilge ar fáil cheana féin ar na taispeántais sna musaeim, sna dánlanna agus sna hionaid oidhreachta. Ós é go bhfuil na suíomhanna sin teoranta go nádúrtha ó thaobh na dtopaicí a chlúdaítear iontu, is áiteanna an-mhaith iad chun tosú ar chorais chainte ríomhbhunaithe a fhorbairt agus a thástáil.

Tá taighde ar bun sa réimse oideachais faoi láthair ina bhfuil úsáid á baint as holagraim agus róbait chun scéalta beatha daoine éagsúla a chaomhnú (féach, mar shampla, Fíor 10.3: duine a tháinig slán ón uileloscadh). D'fhéadfadh sé seo a bheith i bhfad níos éifeachtaí mar mhodh foghlama faoi stair nó faoi theangacha ná díreach cuntas a léamh i leabhair mar go gcuirfeadh sé aos óg an lae inniu go díreach i dteagmháil le samhaltán ó aois atá caite. Faoi mar atá le córais chomhrá ríomhbhunaithe go ginearálta, thógfadh sé seo comhoibriú idir réimsí éagsúla taighde, réaltacht fhíorúil agus staidéar ilmheánach ina measc.

Fíor 10.2: Treoraithe daonna fíorúla i Musaem na hEolaíochta, Bostún (https://www.youtube.com/watch?v=rYF68t4O_Xw)

Fíor 10.3 Comhrá trí hologram: ó 'The Near Future' – David Traum, USC Institute for Creative Technologies (https://www.youtube.com/watch?v=UjI-W1z9ruw)

Gníomhaire teagaisc: Is é atá i gceist le gníomhaire teagaisc ná carachtar ar ríomhaire a cruthaíodh chun tasc áirithe a mhúineadh (Gulz, Haake, Silvervarg, Sjödén, & Veletsianos, 2011). Tá sé i gceist go mbeadh 'intleacht' áirithe ag an gcarachtar agus go mbeadh an cumas ann feidhmiú as a stuaim féin. Murab ionann agus na córais teagaisc sheanbhunaithe, glactar leis de ghnáth go mbíonn foirm chorpartha ag an ngníomhaire maille le híomhá bheo.

10.2.2 Uaillmhianta don Ghaeilge

Cé go mbaineann cuid mhaith de na feidhmeanna atá forbartha go dtí seo le cúrsaí oideachais tá córais chomhrá ríomhbhunaithe ag éirí níos coitianta sa ghnáthshaol lasmuigh den scolaíocht agus is é ár mian ná go mbeadh córais don Ghaeilge againn a bheadh ionchurtha leo siúd atá á bhforbairt do na mórtheangacha domhanda – feidhmeanna a bheadh in úsáid san iliomad suíomhanna.

Chomh fada is a bhaineann le cúrsaí oideachais is é ár n-uaillmhian don Ghaeilge ná go gcruthóimis teagascóir cliste croschineálach a bheadh in ann dul i mbun comhrá le foghlaimeoir. Go hidéalach, bheifí ag iarraidh gnáthchomhrá a nascadh le feidhm oideolaíoch an teagaisc chliste chun tumthimpeallacht a chruthú ina mbeadh an foglaimeoir gafa leis an idirphlé agus go mbraithfeadh sé go raibh ceangal pearsanta aige leis an bhfiontar.

10.3 A bhfuil déanta i leith na Gaeilge go dtí seo

Réimse an-nua é seo don Ghaeilge. Mar sin féin, tá córas bunúsach comhrá, Taidhgín, tógtha don Ghaeilge sa tSaotharlann Foghraíochta agus Urlabhra, Coláiste na Tríonóide, Baile Átha Cliath (Ní Chiaráin & Ní Chasaide, 2016a, 2016b).

Taidhgín mar chéile comhrá: tá Taidhgín tógtha mar mhoncaí beochana agus is féidir leis comhrá a dhéanamh le húsáideoir laistigh de réimse áirithe topaicí. Úsáideann Taidhgín guth sintéiseach ABAIR agus roghnaíonn

an t-úsáideoir pé guth is fearr leis dó féin. Tá Taidhgín fós ina réamhshamhail ach is é an aidhm atá ann é a fhorbairt mar chóras cliste teagaisc teanga, gníomhaire teagaisc agus comhpháirtí comhrá, is é sin na gnéithe ar fad atá inmhianaithe d'fhorbairtí sa Ghaeilge a chorprú ann.

Fíor 10.4: Taidhgín, gníomhaire teagaisc le caint shóisialta agus é i mbun comhrá.
Tabhair faoi deara go ndéantar taifead ar gach comhrá chun anailís a dhéanamh air níos déanaí.

Ós rud é nach raibh aithint cainte don Ghaeilge ar fáil nuair a tógadh é, is trí théacs a dhéantar an t-ionchur agus léitear an t-ionchur seo amach os ard. Freagraíonn Taidhgín i bhfoirm labhartha agus tugtar an téacs scríofa don fhoghlaimeoir chomh maith. Tá na léirmheasanna ar Thaidhgín dearfach go dtí seo agus taispeánann siad gur féidir é a úsáid chun foghlaimeoirí a theagasc le teanga nádúrtha ríomhbhunaithe.

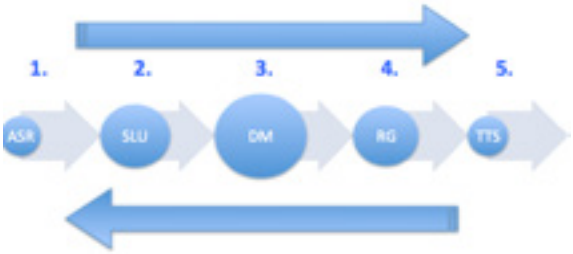
Taidhgín mar Theagascóir Cliste: tá leaganacha eile teoranta de Thaidhgín á bhforbairt mar theagascóirí cliste chun litriú agus pointí gramadaí a mhúineadh. Is i bhfoirm cluichí agus quizeanna a chuirtear na feidhmeanna seo i láthair – Trialacha Taidhgín. Samhlaítear go ndíreoidh forbairtí Thrialacha Thaidhgín ar chluichí a dhearadh agus feidhm ábaint as an staidéar teangeolaíochta a rinneadh go dtí seo (féach Caibidil 1 m.sh. uimhreacha/séimhiú) chomh maith le teicneolaíochtaí teanga idir shean agus nua (féach Caibidil 3, m.sh. WordNets).

Féachtar ar Thaidhgín agus tuilleadh gníomhairí nach é mar bhunábhair atá lárnach d'fhoghlaim ríomhchuidithe teangacha (féach Caibidil 14). Cé nach bhfuil sé ach ina thús, amach anseo táthar ag súil go mbeidh Taidhgín in ann:

- nasc sóisialta a dhéanamh leis an bhfoghlaimeoir – is é sin comhrá a dhéanamh leis an bhfoghlaimeoir i dteanga nádúrtha chun go mbainfidh sé taitneamh nó pléisiúr as an taithí
- teagasc cliste a dhéanamh i réimsí ar leith (m.sh. fuaimniú, gramadach, stór focal, srl. nó chun ábhair eile amhail an eolaíocht a mhúineadh trí Ghaeilge)
- a bheith freagrúil do riachtanais ar leith a bheadh ag foghlaimeoirí: a bheith in ann gnáthbhotúin leanúnacha a dhéanann foghlaimeoirí a aithint agus aiseolas pearsanta a thabhairt má lorgaítear a leithéid

10.4 Conas a oibríonn sé?

Léirítear i bhFíor 10.5 thíos na bunchéimeanna a bhaineann le córas comhrá ríomhbhunaithe a chur ag feidhmiú.



Fíor 10.5: An struchtúr bunúsach atá ag córas comhrá ríomhbhunaithe

Is rud casta é córas comhrá ríomhbhunaithe a thógáil mar bíonn go leor teicneolaíochtaí éagsúla ag teastáil chun teanga daonna a phróiseáil, gach ceann acu ina réimse staidéir ann féin (tá cur síos déanta cheana ar chuid acu i gcaibidil eile den Phlean Digiteach seo) Tá na cúig chéim atá léirithe i bhFig 10.5 luaite anseo thíos agus cur síos déanta orthu:

1. ASR (aithint cainte) ar an ionchur cainte
2. SLU (tuiscint ar an teanga labhartha) chun ciall a bhaint as an gcaint
3. DM (bainistíocht comhrá). Tá sé seo lárnach. Rialaíonn sé na comhpháirteanna atá i gcóras comhrá labhartha agus déanann sé iarracht an freagra is oiriúnaí a fháil ('Gníomh')
4. RG (Giniúint freagra). Cumann sé freagra i dteanga nádúrtha
5. TTS (Sintéis téacs-go-hurlabhra). Iompaíonn sé an téacs isteach ina ghuth don aschur

Labhraíonn an t-úsáideoir isteach sa chóras agus athraíonn sé an chaint go téacs (1). Baintear ciall as cuid de (2). Níor ghá go mbeadh aon rud anseo ach aitheantas focal (mar atá sa leagan reatha de Thaidhgín) nó d'fhéadfadh go mbeadh an anailís i bhfad níos compléascaí ná sin ar bun. Tá 'bainistiú comhrá' i gcroílár an chórais agus déanann sé meaitseáil (i gcóras simplí) nó déanann sé léamh níos casta air má bhíonn an córas féin sách casta. Is as sin a éiríonn an gníomh atá ag teastáil (3), is é sin an freagra is oiriúnaí agus an ceann is fearr a cheapadh chun an t-úsáideoir a spreagadh chun ionchur eile a dhéanamh agus an próiseas a thosú arís (4). An chéad chéim eile ná an freagra a chur in oiriúnt do chomhthéacs an chomhrá suas go dtí sin (m.sh. an aimsir cheart den bhriathar srl.). An chéim dheiridh ná an téacs a aistriú isteach i bhfoirm labhartha don aschur (5).

Úsáideann córais chomhrá ríomhbhunaithe go leor de na bunteicneolaíochtaí agus na hacmhainní teangeolaíochta i dteannta a chéile. Tá cuid acu seo ar fáil cheana féin, cosúil le Céim (1) sintéis chainte (Caibidil 8). Tá an chuma air go bhfuil an taighde atá ar bun chun mothúcháin a chur in iúl trí na guthanna (Caibidil 7), chun sinn a thabhairt níos gaire

do chomhrá nádúrtha. Tá Céim (5) aithint cainte (Caibidil 9) tábhachtach chomh maith ach níl an obair air seo ach ina tús. Mar a dúradh thuas, is féidir le sainchórais aitheanta cainte comhthéascanna a chruthú ina mbeadh Taidhgín níos éifeachtaí mar áis teagaisc. Beidh teacht ar chéimeanna (2) agus (4) trí thaighde ar Phróiseáil Teanga Nádúrtha mar atá léirithe i gCaibidlí 4 agus 5 agus taighde ar Ghiniúint Teanga mar atá leagtha amach i gCaibidil 6. Is é an bainisteoir comhrá croílár an chórais (Céim (3)). Is é seo bunábhar an taighde atá riachtanach agus ansin an nascadh idir é seo agus na heilimintí eile atá luaite thuas.

Cur Chuige Riailbhunaithe v Meaisínfhoghlaim
Tá dhá bhealach éagsúla ann chun dul i mbun córais chomhrá ríomhbhunaithe a fhorbairt (féach an plé i gCaibidil 1.7). Beidh gá le córas riailbhunaithe ag an tús atá bunaithe ar theangeolaíocht, tuiscint ar chomhthéacs sa saol réadúil, tuiscint ar mhúnlaí de phrosóid na cainte a ghabhann le comhrá chomh maith le corpais chomhrá atá anótáilte. Ag pointe áirithe is féidir dul i mbun straitéisí meaisínfhoghlama ach ní mór a bheith cáiréiseach faoi seo i gcás na Gaeilge: ar éigean a éireoidh le córas atá bunaithe ar mheaisínfhoghlaim sa Bhéarla (féach Caibidil 1).

10.5 Moltaí: cur chuige céimnithe

Beidh gá le pleanáil chúramach scóipiúil chun córais chomhrá ríomhbhunaithe a fhorbairt. Tá sé tábhachtach go leagfaí síos príomhspriocanna sainiúla agus na córais á dtógáil – feictear gur feidhmeanna oideachasúla agus feidhmeanna rochtana do dhaoine faoi mhíchumas is mó atá i gceist i gcás na Gaeilge. Seo thíos na céimeanna a bheidh i gceist chun córais chomhrá ríomhbhunaithe a sholáthair:

1. Córas bainistíochta comhrá a fhorbairt a bheidh mar thaca le forbairt níos soifisticiúla a dhéanamh ar an leagan de Taidhgín atá anois ann nó ar chinn eile nach é. Ní mór iniúchadh a dhéanamh ar pháirtnéirí éagsúla comhrá, m.sh.
   a. teagascóir cliste
   b. gníomhaire teagaisc
   c. gníomhaire comhrá
   d. tum-ardán idirghníomhach
2. Taighde ar chaint idirghníomhach lena léirítear mothúcháin ina mbeidh, mar shampla:
   (a) corpais chomhrá don Ghaeilge á mbailiú agus á bpróiseáil. I bprionsabal, ba cheart go ndéanfaí é seo i gcomhpháirt le corpas cainte a bhailiú don tsintéis (Caibidil 8) agus don aithint cainte (Caibidil 9).
   (b) taighde ar chaint eispriseach, i gcomhpháirt leis an obair atá luaite i gCaibidlí 2, 4 agus 7. Tá sé i gceist anseo go mbeadh taighde leanúnach ann ar chaint na ndaoine agus ar an tuiscint a bhaineann éisteoirí as na gnéithe éagsúla de na mothúcháin agus eile atá ar fáil sa chaint sin.
3. Forbairt córais chomhtháite: áit ar féidir le háiseanna éagsúla cainte agus próiseáil nádúrtha teanga feidhmiú i dteannta a chéile ar shlí a chuirfidh le forbairt chórasach áiseanna don fhíorshaol.
4. Ábhair theagaisc agus suíomhanna foghlama a fhorbairt: beidh gá le comhoibriú agus ionchur ó shaineolaithe oideachais.
5. Comhoibriú a chothú le grúpaí atá ag obair i réimsí na réaltachta fíorúla /réimsí ilmhódacha eile nach iad: bheadh gá le comhthionscadail ionas go bhféadfaí ardáin a fhorbairt go pointe ina mbeadh carachtair a bhfuil pearsantachtaí éagsúla acu, léiriúcháin éagsúla, agus timpeallachtaí réaltachta fíorúla, mar shampla, ar fáil.
6. Clár chun feidhmchláir don fhíorshaol a fhorbairt a thagann i dtír ar na teicneolaíochtaí agus na háiseanna is nua-aimseartha. Beidh gá le córas lena spreagfar tionscadail atá forbartha go leibhéal an chruthúnais coincheapa, ionas go bhfásfaidh isteach ina n-ardáin seasmhacha, ar féidir iad a chur os comhair an phobail.
7. Measúnú leanúnach le spriocúsáideoirí a thabharfaidh treoir do gach céim den fhorbairt, lena n-áirítear na túschéimeanna dearaidh.

Mar chonclúid, samhlaítear fáinne atriallach ina n-úsáidfear an saineolas agus na háiseanna atá luaite i Roinn I i dteannta na mbunteicneolaíochtaí atá luaite i Roinn II. De réir mar a thiocfaidh córais chomhrá chun tosaigh beidh siad lárnach i gcuid mhaith de na feidhmeanna atá luaite i Roinn III, agus d'Foghlaim Ríomhchuidithe Teangacha (Caibidil 14) go háirithe.

Tagairtí

1. Gulz, A., Haake, M., Silvervarg, A., Sjödén, B., & Veletsianos, G. (2011). Building a Social Conversational Pedagogical Agent : Design Challenges and Methodological approaches. In D. Perez-Marin & I. Pascual-Nieto (Eds.), Conversational Agents and Natural Language Interaction: Techniques and Effective Practices (pp. 128–155). Information Science Reference. http://doi.org/10.4018/978-1-60960-617-6.ch006

2. McTear, M. (2002). Spoken Dialogue Technology: Enabling the Conversational User Interface. ACM Computing Surveys, 34(1), 90–169.

3. Ní Chiaráin, N., & Ní Chasaide, A. (2016a). Chatbot Technology with Synthetic Voices in the Acquisition of an Endangered Language: Motivation, Development and Evaluation of a Platform for Irish. In 10th edition of the Language Resources and Evaluation Conference, 23-28 May 2016 (pp. 3429–3435). Portorož (Slovenia).

4. Ní Chiaráin, N., & Ní Chasaide, A. (2016b). Faking Intelligent CALL: the Irish context and the road ahead. In E. Volodina, G. Grigonytė, I. Pilán, K. Nilsson Björkenstam, & L. Borin (Eds.),Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition at SLTC (pp. 60–65). Umeå, Sweden: Linköping University Electronic Press, Linköpings universitet.

Caibidil 11
Meaisínaistriúchán

11.1 Céard is meaisínaistriúchán ann?

Is éard is Meaisínaistriúchán Uathoibríoch ann ná próiseas atá bunaithe ar bhogearraí ina ndéantar inneachar digiteach (téacs, urlabhra, srl.) a chlaochlú ó theanga amháin go teanga eile. Is tasc an-chasta é seo lena mbaineann ciall agus gnásanna teachtaireachta foinsí áirithe a thuiscint agus teachtaireacht choibhéiseach a chruthú ina gcuirtear an chiall chéanna in iúl trí bhíthin gnásanna na sprioctheanga a úsáid.

Tá dul chun cinn á dhéanamh go síoraí seasta ar an teicneolaíocht is bun le córais an mheaisínaistriúcháin (MA) nua-aimseartha. Má tá féin, tá roinnt mórpharaidímí sa réimse, mar atá iad siúd atá bunaithe ar shonraí (MA staitistiúil agus néarach) agus iad siúd atá bunaithe ar rialacha (MA atá bunaithe ar theanga). Tá láidreachtaí agus laigí ag roinnt leis an dá chur chuige araon. Ar na saolta seo, i dtimpeallachtaí taighde agus tionsclaíochta araon, is é an MA néarach an múnla is mó a mbaintear leas as.

Dhá chuspóir atá leis an meaisínaistriúchán. Cuspóir amháin is ea meabhair nó ciall a bhaint as píosa téacs atá i dteanga nach bhfuil eolas ag an úsáideoir uirthi. Baineann an cuspóir eile le cuspóirí réamhaistriúcháin i dtimpeallacht ghairmiúil, ina ndéanann aistritheoir gairmiúil profléamh nó iar-eagarthóireacht ar aistriúchán a fhaightear ó chóras MA. Beidh an léitheoir taithíoch ar an mí-úsáid neamhbheartaithe a bhaintear as saorchórais MA ar líne, ina ndéantar dearmad ar an gcéim iar-eagarthóireachta go minic. Is éard is cúis leis sin, tríd is tríd, gur minic nach soiléirítear don úsáideoir go bhfuil sé bunriachtanach an t-aschur a aistríodh go huathoibríoch a sheiceáil agus a phrofléamh. Chomh maith leis sin, bíonn caighdeán na saorchóras MA ar líne an-éagsúil lena chéile, ó phéire aistriúcháin amháin go péire aistriúcháin eile, agus ní thugtar aon chomhartha sna córais chéanna ar fheabhas nó ar iontaofacht choibhneasta aon phéire aistriúcháin faoi leith thar a chéile.

Sa chaibidil seo féachtar le míniú a thabhairt ar an gcaoi a n-oibríonn an meaisínaistriúchán, an chaoi a bhféadtar é a fheabhsú chun críche na Gaeilge, agus an chaoi a bhféadtar feidhm a bhaint as le costais agus le saothar aistriúcháin a laghdú, nuair a úsáidtear mar chuid de phróiseas aistriúcháin é.

11.2 Cad chuige a bhfuil sé tábhachtach agus cé dó a bhfuil sé tábhachtach?

Tá athrú suntasach tagtha ar chúram an aistritheora le fiche bliain anuas ó tugadh an teicneolaíocht aistriúcháin isteach. D'fhonn cuidiú le hualach oibre an aistritheora a laghdú, d'fhonn bisiúlacht a fheabhsú, agus d'fhonn cuidiú le laghdú a dhéanamh ar an mbearna idir riar agus éileamh, is amhlaidh atá earnáil an aistriúcháin agus na seirbhísí teanga ar a ndícheall ag iarraidh uirlisí níos fearr a chur ar fáil d'aistritheoirí lena gcuid oibre a chur i gcrích. Go stairiúil, is iad gluaiseanna, bunachair sonraí téarmaíochta agus bogearraí Cuimhne Aistriúcháin (CA) a bhí sa mhéid sin. Faoi mar atá feabhas tar éis teacht le deireanas ar bhuneolaíocht, ar bhogearraí, agus bhonneagar an Mheaisínaistriúcháin, is amhlaidh atá an úsáid as uirlisí Meaisínaistriúcháin (MA) ina gnáthchleachtas i dtimpeallacht ghairmiúil aistriúcháin ar fud an domhain. Tá glacadh forleathan leis an úsáid as MA sa riarachán poiblí i dtíortha dátheangacha, amhail Ceanada agus an Bhreatain Bheag.

Is féidir úsáid a bhaint as uirlisí nua-aimseartha MA ar an iomad bealach. De ghnáth, baineann an t-aistritheoir úsáid astu mar uirlis bhreise tacaíochta. Ar an gcaoi sin, cuireann an t-aistritheoir chun oibre mar is gnách, ag úsáid gluaiseanna, bainc sonraí téarmaíochta nó bogearraí CA, ach tá de rogha bhreise aige/aici glacadh le moltaí a bhíonn á nginiúint go huathoibríoch ag an MA chomh maith. Tugann a leithéid de chomhtháthú saoráideach caoi don aistritheoir úsáid a bhaint as an rud is úsáidí agus is éifeachtaí don tasc atá idir lámha agus is féidir leis/léi beartú ar oibriú ó bhonn, ar ghnéithe de na tacaíochtaí éagsúla atá ar fáil a úsáid in éineacht, nó iar-eagarthóireacht a dhéanamh ar mholadh inúsáidte a ghin an t-ineall MA go huathoibríoch.

Le blianta fada, táthar ag taispeáint i staidéir fhairsinge an chaoi a ndéantar bisiúlacht a fheabhsú, de bharr MA (a chuireann leis an úsáid as uirlisí cuimhne aistriúcháin go minic) a chomhtháthú lena leithéid de shreabhadh oibre, sa taighde earnálach agus acadúil araon, .i. (1,2). Chomh maith leis sin, sa chás ina mbíonn aistritheoirí ag plé le téacsanna nua nó téacsanna nach bhfuil siad taithíoch orthu, ina mbíonn líon na meaitseálacha sa chuimhne aistriúcháin íseal, tá léirithe gur mó bisiúlacht na n-aistritheoirí agus iad ag úsáid aschur meaisínaistriúcháin ná a macasamhail agus iad ag aistriú dá uireasa. Ní hé

amháin gur comhartha iad torthaí na staidéar seo ar mhéadú suntasach ar acmhainn táirgthe aistritheoirí a úsáideann uirlisí MA, rud a théann i ngleic leis an mbearna idir riar agus éileamh dá réir, ach is comhartha iad ar na costais shuntasacha a fhéadtar a shábháil chomh maith céanna. Fágann sin gur féidir méid áirithe den obair bhreise atá ag teastáil le freastal ar an éileamh (atá ag dul i méid) a dhéanamh gan a bheith thíos le costas breise oibre tríd an úsáid as uirlisí aistriúcháin nua-aimseartha ina mbaintear leas as MA.

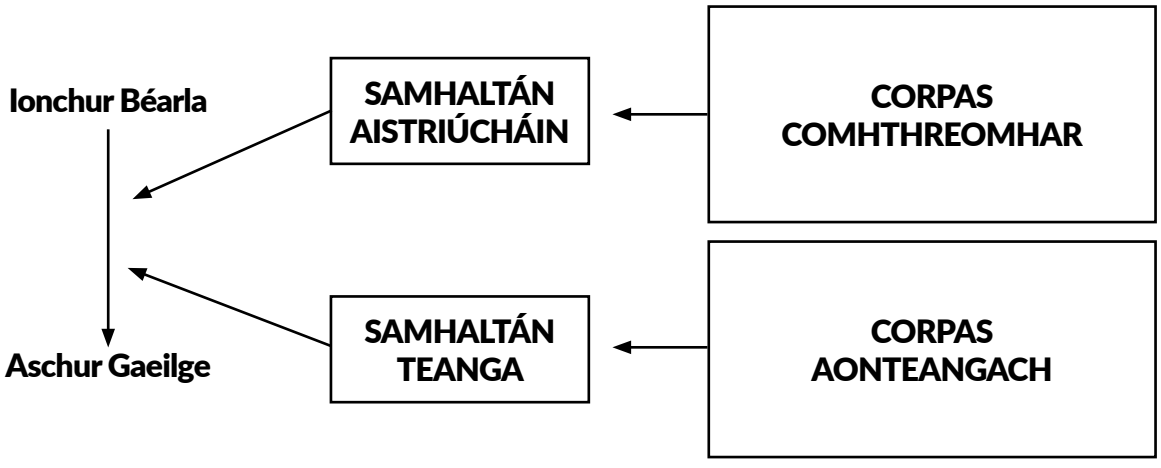
Ina cheann sin, agus is díol suntais an méid seo a leanas i gcás teanga mionlaigh ar nós na Gaeilge, is féidir stádas na teanga a dhaingiú trí bhíthin seirbhísí feabhsaithe teanga a mbeidh méadú ar inneachar digiteach Gaeilge mar thoradh orthu, rud a chosnóidh cearta na saoránach teagmháil a dhéanamh leis an Stát agus úsáid a bhaint as seirbhísí i nGaeilge dá réir.

Ar bhonn náisiúnta, tá éileamh mór ar chórais aistriúcháin Béarla go Gaeilge (EN>GA) i measc sheirbhísí poiblí na hÉireann agus i measc gnólachtaí eile a fheidhmíonn in Éirinn agus a dhéanann freastal ar mhargadh na hÉireann. I dtaca le haistriúcháin reachtúla go háirithe, tá riaráiste oibre tar éis teacht chun cinn ina bhfuil an t-éileamh ar an aistriúchán níos mó ná an riar [3]. I gcomparáid leis an méid ábhar atá le haistriú, tá líon sách beag ann d'aistritheoirí a bhfuil na cáilíochtaí cuí acu agus acmhainn táirgthe sách teoranta acu siúd chomh maith. Faoi láthair, tá an bhearna atá de thoradh ar an gcás sin á leathnú, ó cuireadh deireadh leis an maolú ar reachtaíocht agus ar cháipéisí an Aontais Eorpaigh a aistriú ag deireadh 2021. Mar a bhíonn amhlaidh i gcás gach teanga Eorpaí, beidh tionchar suntasach ag an teicneolaíocht aistriúcháin i dtaca leis an mbearna seo a laghdú ar bhonn náisiúnta agus ar bhonn Eorpach araon.

11.3 Conas a oibríonn an meaisínaistriúchán?

Tá an dearadh atá ar a lán de na córais MA 'faoi thiomáint sonraí'. Cuireann innealtóirí ríomhaireachta sonraí comhthreomhara (.i. téacsanna a aistríodh go gairmiúil cheana) isteach sa chóras, ionann is go gcuireann siad traenáil air, agus uaidh sin foghlaimíonn an córas an chaoi le haistriúchán nua a thuar (féach Fíor 11.1 mar a léirítear an chaoi a n-oibríonn MA Staitistiúil). Déanann algartam staitistiúil amach na comhfhreagairtí idir na teaghráin litreacha sa dá theanga. Ina dhiaidh sin, i bhfianaise a mhinice a aimsíonn an t-algartam na comhfhreagairtí sin, déanann an t-algartam scór dóchúlachta a cheapadh, scór a mbaintear úsáid as leis an aistriúchán is mó dóchúlacht ar theaghrán ionchuir faoi leith a dhéanamh amach. Dá mhéad na samplaí d'aistriúcháin a aimsíonn an córas, is ea is fearr na tuartha agus is ea is mó an cruinneas. Cuirtear feabhas ar an aistriúchán a gineadh le samhail teanga, samhail a bhfuil traenáil i dtéacs aonteangach de chuid na bunteanga uirthi. Ar an lámh eile de, tá an meaisínaistriúchán néarach bunaithe ar an meaisínfhoghlaim agus ar líonraí néaracha [7-9]. Déantar na sonraí comhthreomhara traenála a ghlacadh agus na focail a choinbhéartú go léiriúcháin veicteora (ionchóduithe matamaiticiúla), sula dtugtar faomhadh do léiriúchán le haghaidh na habairte iomláine trí líonra néarach. Ar an gcaoi sin, is fearr i bhfad mar a thuigtear agus a mhínítear comhthéacs focail. Córais 'ceann go ceann' is ea a thugtar ar chórais MAN mar nach bhfuil iontu siúd na comhpháirteanna ar leithligh a bhíonn i gcórais MAS.

Tá de bhuntáiste ag na modhanna sin atá faoi thiomáint sonraí gur féidir iad a chur i bhfeidhm go tapa, go bhfuil siad láidir nuair atáthar ag déileáil le drochshonraí/sonraí torannacha, agus gur féidir iad a thiúnáil chun téacs i réimse sainiúil a aistriú (.i. an leigheas, an dlí, an riarachán poiblí). Mar sin féin, tá an dá chur chuige ag brath ar an tuiscint go bhfuil teacht ar shonraí traenála feiliúnacha agus go bhfuil teacht ar a ndóthain de na sonraí céanna (bíonn níos mó sonraí de dhíth ar MAN ná macasamhail MAS), rud a bhíonn doiligh teacht air go minic i gcás na Gaeilge de.



Fíor 11.1 An leagan amach a bhíonn ar chóras bunúsach Meaisínaistriúcháin Staitistiúil

Féadann sé a bheith dúshlánach teacht ar shonraí maithe comhthreomhara, go háirithe i gcás teangacha nach gcuirtear an oiread sin acmhainní ar fáil dóibh, ar nós na Gaeilge. Go minic, is é an príomhthasc a bhíonn i gceist le corpas comhthreomhar a chruthú ná sonraí a aistríodh cheana féin a aithint agus a bhailiú ó aistritheoirí nó ó chomhlachtaí aistriúcháin, agus teagmháil a dhéanamh leis na ranna rialtais cuí nó leis na heagraíochtaí leathstáit cuí ag iarraidh a n-aistriúchán oidhreachta. Anuas air sin, úsáidtear uirlisí bogearraí a dtugtar 'ransaitheoirí gréasáin' orthu le teacht ar théacsanna comhthreomhara ar an idirlíon. Tá sé ríthábhachtach deimhin a dhéanamh de go bhfuil ailíniú ceart ar abairtí d'fhonn an córas MA a bhacadh ar aistriúcháin mhíchearta a fhoghlaim. Cé go bhfuil MAS agus MAN an-mhaith chuig friotal dúshlánach nathach a láimhseáil, ba chóir cuimhneamh air freisin gur féidir leis earráidí bunúsacha gramadaí agus loighce a dhéanamh d'uireasa comhpháirt atá bunaithe ar rialacha.

Ar an lámh eile de, úsáideann córas MA atá bunaithe ar rialacha bunrialacha teanga agus bunrialtachtaí teanga le haistriúchán a chruthú. Is éard a bhíonn sa mhodh sin ná ríomhaire a bheith á ionchódú le heolas daonna ar ghramadach agus ar struchtúr na teanga foinsí ag innealtóirí teanga, agus déanann siadsan an t-eolas sin a mhapáil ar an sprioctheanga ina dhiaidh sin. Úsáidtear an fhaisnéis sin leis an aschur deiridh aistriúcháin a ghiniúint trí bhíthin leas a bhaint as eolas faoi ghramadach agus faoi struchtúr na sprioctheanga. Is snasta i bhfad an cur chuige sin ar bhealaí go leor agus is féidir leis miondifríochtaí teanga a bhíonn deacair a bhaint as na sonraí iad féin a chur san áireamh. Mar sin féin, braitheann sé ar fháil a bheith ar uirlisí do gach teanga ina bhféadtar próiseáil teanga a dhéanamh ar shonraí, agus ar obair láimhe an-sainiúil le hionchódú a dhéanamh ar na rialacha éagsúla atá ag teastáil chomh maith céanna. Fágann sin gur faide an t-achar ama a thógann sé le córais ina n-úsáidtear MA atá bunaithe ar rialacha a chur i bhfeidhm agus nach bhfuil sé chomh héasca sin iad a choinbhéartú ó phéire teangacha amháin go péire teangacha eile.

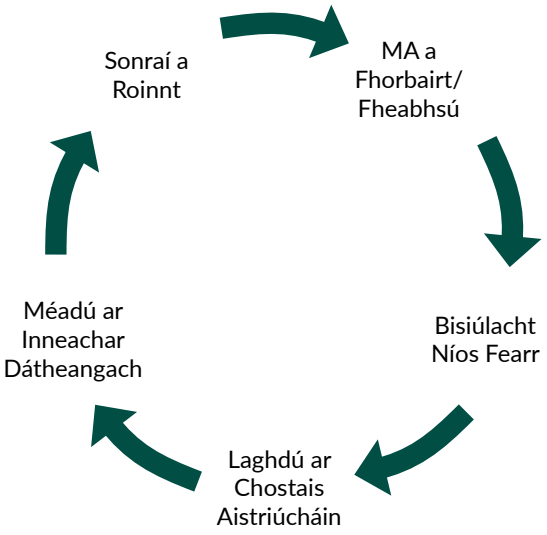
I gcás an dá pharaidím araon, tá modhanna tagartha ann i dtaca le cur i bhfeidhm de arb iad na príomh-mhodhanna iad i measc lucht taighde agus earnála araon. Tá an dá fhoireann uirlisí ar fáil trí na hardáin foinse oscailte mar atá Apertium , Moses agus OpenNMT le haghaidh MA atá bunaithe ar rialacha, MA atá bunaithe ar staitisticí agus MA néarach faoi seach. Tá torthaí dearfacha i dtaca le hinnill MA a fhorbairt le haghaidh aistriúchán EN>GA] á mbaint as obair atá ar siúl faoi láthair ag grúpaí taighde in Éirinn atá ag úsáid na n-ardán sin.

11.4 A bhfuil déanta i leith na Gaeilge go dtí seo

Le blianta beaga anuas, rinneadh roinnt staidéar ar an meaisínaistriúchán le haghaidh na Gaeilge. Ó thaobh an chórais staitistiúil de, tá taighdeoirí in Ollscoil Chathair Bhaile Átha Cliath tar éis córas MA Moses-bhunaithe a fhorbairt chun go n-úsáidfidh aistritheoirí inmheánacha sa Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meán (RTCEGSM) é mar chuid dá sreabhadh oibre aistriúcháin [4]. Is féidir an córas sin, ar a dtugtar 'an Tapadóir', a úsáid chun an téacs a réamhaistriú sa chaoi agus nach mbeadh le déanamh ag aistritheoirí gairmiúla ach ceartú (iar-eagarthóireacht) a dhéanamh ar an aistriúchán a mhol an córas. Tá an córas ag teacht le caighdeán a aithnítear ina chaighdeán feiliúnach do shuíomh gairmiúil iar-eagarthóireachta, de bharr gur cuireadh traenáil shainiúil air le haistriúcháin a rinneadh cheana agus atá i seilbh na Roinne.

Tá staidéir ar mhodhanna éagsúla le MA Gaeilge a fheabhsú á ndéanamh i gcónaí in Ollscoil Chathair Bhaile Átha Cliath mar chuid de staidéir iarchéime. Tá measúnú déanta ag saineolaithe an Mheaisínaistriúcháin ar úsáid an Mheaisínaistriúcháin Néaraigh le haghaidh na Gaeilge, agus d'aithin siadsan bealaí le halgartaim a fheabhsú d'fhonn torthaí níos fearr a bhaint amach, in ainneoin go bhfuil siad ag plé le sonraí teoranta traenála [10]. Tá modúil teanga á gcomhthathú leis na córais sin le cinntiú go ndéanfar an saothar iar-eagarthóireachta a bhíonn ar aistritheoirí a laghdú a oiread agus is féidir [6]. Rinneadh taighde ar MAS agus ar MAN don Ghaeilge in Ionad Taighde Insight in Ollscoil na Gaillimhe freisin [11-12].

Sa taighde ar an MA le haghaidh na Gaeilge a rinneadh go dtí seo i gColáiste na Tríonóide, Baile Átha Cliath, díríodh go príomha ar na gnéithe teanga a bhaineann leis an aistriú ó Bhéarla go Gaeilge. Le blianta beaga anuas, rinneadh roinnt staidéar [13-16] ina ndearnadh iniúchadh ar ghnéithe a bhaineann le húsáid córas MA atá bunaithe ar rialacha amhail Apertium [5] in éineacht le teicneolaíocht atá bunaithe ar rialacha le haghaidh na Gaeilge [17], atá le húsáid i réitigh hibrideacha MA. I staidéir ina ndéantar iniúchadh ar an úsáid a bhaintear as faisnéis teanga i gcórais Meaisínaistriúcháin Néaraigh, baineadh amach torthaí a bhfuil gealladh fúthu agus is dócha gur ábhar torthúil taighde a bheidh ann le haghaidh an phéire Béarla-Gaeilge.



Fíor 11.2 An dea-chleachtas maidir le sonraí a chomhroinnt d'fhonn córais mheaisínaistriúcháin a fheabhsú

De bhrí go mbraitheann córais ardchaighdeáin MA atá bunaithe ar shonraí ar theacht a bheith acu ar shonraí teanga atá ar ardchaighdeán (idir chomhthreomhar agus aonteangach), is fiú na hiarrachtaí i dtaca lena leithéid de shonraí a bhailiú agus a chomhordú atá ar bun ar fud na tíre a lua. I mí na Bealtaine 2019, trí thionscadal an Bhonneagair Eorpaigh Acmhainní Teanga (ELRI), sheol taighdeoirí in Ollscoil Chathair Bhaile Átha Cliath Stáisiún Athsheachadta Náisiúnta (SAN) na hÉireann. Tríd an Stáisiún Athsheachadta – a athsheoladh le déanaí faoin ainm 'eStór' ¬– bailítear sonraí Gaeilge – is féidir le húsáideoirí a gcuid sonraí dátheangacha féin agus a gcuid téarmaíochtaí féin a uaslódáil ar an tairseach, agus i gcás baisceanna de chomhaid chomhthreomhara a uaslódáil, is amhlaidh a chruthófar comhaid cuimhne aistriúcháin go huathoibríoch don úsáideoir. Is féidir le haon duine atá ag obair leis an nGaeilge sa riarachán poiblí iarratas a dhéanamh ar chuntas le haghaidh na tairsí. Ní hamháin go n-úsáidfear gach sonra a bhaileofar lena roinnt taobh istigh de chóras eTranslation an Choimisiúin Eorpaigh (MA Néarach), ach aon chomhlacht riaracháin phoiblí de chuid na hÉireann ar mhaith leis rochtain a fháil ar na comhaid cuimhne aistriúcháin nó atá ag iarraidh a chóras MT saincheaptha féin a fhorbairt, féadann seisean na sonraí a úsáid chomh maith. Cuir i gcás, de bharr thionscadal PRINCIPLE a bhí á mhaoiniú ag an Aontas Eorpach, chonacthas córais MA shaincheaptha á bhforbairt agus á n-úsáid ag aistritheoirí i Rannóg an Aistriúcháin, i bhForas na Gaeilge agus in Ollscoil na Gaillimhe. I bhFíor 11.2, léirítear an tairbhe fhadtéarmach a bhaineann le sonraí a chomhroinnt i gcás úsáideoirí féideartha de.

11.5 Moltaí

Is gá ard-bhogearraí agus bogearraí nua-aimseartha in éineacht a fhorbairt go leanúnach le cinntiú go mbeidh teacht ar bhonneagar meaisínaistriúcháin Béarla-Gaeilge atá ar ardchaighdeán. De bharr na hoibre atá déanta sa réimse seo cheana féin, táthar tar éis inmharthanacht agus tionchar dearfach na teicneolaíochta seo agus an chur chuige seo a léiriú. Mar sin féin, níl an méid sin cuimsitheach a dhóthain go fóill ná níl glacadh forleathan leis in earnálacha go leor ach an oiread. D'fhonn tacú le dul chun cinn leanúnach sa réimse seo, aithníodh cúig ghníomh mar réimsí ar ghá aire a thabhairt orthu láithreach bonn. Tabharfaidh na gníomhartha sin go léir caoi don phobal reatha aistriúcháin, agus d'fhoirne taighde agus forbartha, idir acadúil agus tionsclaíoch, leanúint orthu agus nuáil a dhéanamh sa réimse seo d'fhonn na huirlisí agus na hacmhainní atá ag teastáil a sholáthar le cinntiú go mbeidh na huirlisí agus na hacmhainní aistriúcháin is cothroime le dáta ag tacú go cuí lenár dteanga.

11.5.1 Sonraí a Bhailiú agus Chomhroinnt

Cibé cé acu le traenáil a chur ar chórais staitistiúla nó le comhpháirteanna próiseála teanga a bhailíochtú agus a thástáil a bhaintear feidhm as corpais dhigiteacha chomhthreomhara, bíonn na corpais chéanna ríluachmhar le huirlisí MA a fhorbairt (féach Caibidil 3 maidir le Corpais). Ar fud na tíre, níl dóthain úsáide á baint as sonraí teanga luachmhara faoi mar atáthar á meas faoina luach freisin. Tá teacht go héasca ar a leithéid de shonraí mar thoradh ar ábhair Ghaeilge nó ar obair aistriúcháin Béarla-Gaeilge a bheith á bhfoilsiú. Níl le déanamh ach na sonraí sin a bhailiú, a stóráil go cuí agus a chur ar fáil lena n-athúsáid faoi cheadúnas ceadaitheach cuí (.i. amhail CC-BY , arb é an ceadúnas é arna roghnú chun athúsáid a bhaint as sonraí poiblí faoi Threoir AE maidir le Sonraí Oscailte).

Ba chóir go mbeadh na sonraí sin i bhformáid fheiliúnach ar féidir le meaisín í a léamh amhail XML, TMX, XLIFF, srl (nó i gceann ar bith d'fhormáidí comharbasacha/díorthacha na bhformáidí céanna), rud a mhéadófar ar chumas an duine na sonraí a athúsáid agus a chur i bhfeiliúint an athuair go huathoibríoch gan gá a bheith le hailíniú duainéiseach láimhe. Nuair is féidir, ba chóir go gcomhroinnfí na sonraí sin gan srian trí thairsí amhail an Stáisiún Athsheachadta Náisiúnta/ eSTÓR ar mhaithe le leas an phobail agus ba chóir go molfaí cultúr i dtaca le sonraí a chomhroinnt i measc na ngeallsealbhóirí aistriúcháin, rud a chuideoidh le dea-chleachtas a chothú maidir le sonraí a sholáthar agus a athúsáid. Tá sé tábhachtach freisin go gcuirfear oideachas ar shealbhóirí sonraí maidir le luach a gcuid sonraí teanga. Ba chóir comhairle a chur ar fáil

i dtaca leis an gcaoi is fearr le conarthaí aistriúcháin a dhréachtú le gnólachtaí aistriúcháin tríú páirtí chun deimhin a dhéanamh de, nuair is féidir, go gcuirtear comhaid cuimhne aistriúcháin nó a leithéid d'fhothairgí ar ais chuig an eagraíocht nó chuig an roinn státmhaoinithe ag a bhfuil an t-iarratas aistriúcháin á mhaoiniú. Is féidir na cuimhní aistriúcháin sin a úsáid le costais chothroma aistriúcháin a idirbheartú amach anseo, agus le sonraí traenála le haghaidh córais MA náisiúnta a chur ar fáil freisin.

Tacófar lena leithéid de chleachtais bhainistíochta sonraí trí bheartas an Aonaid um Shonraí Oscailte maidir le rochtain dhigiteach ar shonraí poiblí na hÉireann a sholáthar agus ba chóir go gcloífeadh na daoine siúd a oibríonn le sonraí Gaeilge sna comhlachtaí ábhartha leis an mbeartas sin.

11.5.2 Idirphlé a dhéanamh le tionscnaimh Eorpacha níos leithne

Nuair is féidir, ba chóir aon fhorbairt bhreise sa réimse seo a dhéanamh i gcomhar le tionscnaimh níos leithne sa réimse féin, nó ar bhealach a léiríonn feasacht ar na tionscnaimh chéanna. Tá iarrachtaí pobail agus AE ar siúl cheana féin i dtaca le sonraí agus le huirlisí teanga a bhailiú agus a chomhroinnt, e.g., META-SHARE, CLARIN, ELRC, ELRI, European Language Grid (ELG), iarrachtaí ar féidir úsáid a bhaint astu le leas na Gaeilge a dhéanamh.

Is den riachtanas é cothabháil leanúnach a dhéanamh ar an Stáisiún Athsheachadta Náisiúnta/eSTÓR chomh maith le tacaíocht leanúnach a thabhairt dó d'fhonn bainistiú lárnach ar shonraí náisiúnta aistriúcháin a éascu, agus d'fhonn cinntiú go mbainfear athúsáid agus giaráil as acmhainní atá ann cheana féin nuair is gá. Ba chóir go mbeadh an tairseach seo ceangailte le Tairseach Shonraí Oscailte na hÉireann agus ba chóir go n-éascófar comhtháthú le haon chóras lárnaithe a bheidh ann amach anseo, amhail an tSeirbhís Aistriúcháin Chomhroinnte.

Ar bhonn Eorpach, táthar ag glacadh leis an gcur chuige seo i leith traenáil a chur ar chórais MA atá faoi thiomáint sonraí (eTranslation) agus tá fáil orthu cheana féin acu siúd sa riarachán poiblí ar fud na hEorpa. Mar sin féin, níl an leagan Béarla-Gaeilge de chóras eTranslation sásúil i gcomparáid le formhór theangacha eile an Aontais. Is éard is príomhchúis leis sin a laghad sonraí aistriúcháin Béarla-Gaeilge a bhí ar fáil le traenáil a chur ar an gcóras le linn an mhaolaithe.

Tá an Coimisiún Eorpach tar éis sraith treoirlínte a mholadh den chéad uair – treoirlínte ina gcuimsítear na cuspóirí agus na bonneagair seirbhísí digiteacha éagsúla lena n-áirítear an t-aistriúchán uathoibríoch – CEF.AT agus an Clár don Eoraip Dhigiteach. Ar bhonn náisiúnta, bhí iarrachtaí á ndéanamh cheana leis an teacht atá ar shonraí traenála i gcomhair eTranslation a mhéadú tríd an tSaoráid um Chónascadh na hEorpa (SCE) trí thionscadail ELRI agus PRINCIPLE. Bhí ról lárnach ag rannpháirtíocht na hÉireann in European Language Resource Coordination (ELRC) i dtaca leis an dul chun cinn atá déanta ar MA na Gaeilge. Comhlachtaí riaracháin de chuid na hÉireann a bhfuil maoirseacht á déanamh acu ar dhul chun cinn an aistriúcháin agus na teicneolaíochta aistriúcháin, beidh orthu deimhin a dhéanamh de go mbeidh ailíniú le gníomhaíochtaí agus cláir Eorpacha ann, mar go bhfuil go leor tairbhe le baint as comhoibriú le ballstáit eile i dtaca leis an dul chun cinn atá siadsan a dhéanamh ar theicnící agus ar bhonneagair.

11.5.3 Taighde agus forbairt bhreise ar theicneolaíocht an mheaisínaistriúcháin le haghaidh na Gaeilge

Ní féidir dán MA na Gaeilge a fhágáil faoi chúram gnólachtaí domhanda teicneolaíochta amhail Google nó Microsoft. Thar aon ní eile, ba chóir go gcuirfí i gcrích riachtanais ghairmiúla aistriúcháin mar a bhaineann leis an riarachán poiblí, amhail eTranslation nó córais MA shaincheaptha a bhfuil tiúnáil shainiúil orthu go mbeidh siad i bhfeiliúint do sheanrá an téacs atá ag teastáil, i dtimpeallachtaí atá slán. Ó thaobh fhorbairt na gcóras Gaeilge le haghaidh eTranslation, bheadh tairbhe le baint as na tionscadail chomhair atá ar siúl cheana féin le hArd-Stiúrthóireacht an Aistriúcháin sa Choimisiún Eorpach. Is féidir córais MA shaincheaptha a fhorbairt trí thacaíocht ollscoileanna na hÉireann nó trí ghrúpaí taighde.

Sa taighde agus san fhorbairt tosaigh a luadh thuas, táthar tar éis a thaispeáint go bhfuil an fhéidearthacht ann teicneolaíochtaí meaisínaistriúcháin a fhorbairt le haghaidh na Gaeilge. Is léir cheana féin ar na tionscadail ghearrthéarmacha tosaigh sin an luach a bhaineann le hinfheistíocht a dhéanamh i dtionscadail agus i gcláir thaighde agus forbartha a thacaíonn le saothar MA. Anuas air sin, ba chóir go ndéanfaí an taighde sin a leathadh ionas go gcuimseofaí ann an t-aistriúchán urlabhra go hurlabhra, réimse atá ag teacht chun cinn agus ina dtugtar le chéile disciplíní na teicneolaíochta urlabhra agus an mheaisínaistriúcháin.

Tá tacaíocht bhreise do chláir agus do thionscadail fhadthéarmacha dá leithéid ríthábhachtach d'fhonn go ndéanfar an dul chun cinn sin agus d'fhonn go ndéanfar na teicneolaíochtaí seo a thaighde agus a fhorbairt mar is ceart le gur féidir iad a úsáid.

11.5.4 Measúnú daonna ar aschur meaisínaistriúcháin le haghaidh na Gaeilge

Is minic a úsáidtear méadraigh uathoibríocha (.i.BLEU (20) and Meteor (21)) mar threoirlínte le comhartha a thabhairt ar chaighdeán agus ar iontaofacht an aschuir aistriúcháin ó chóras meaisínaistriúcháin. Mar sin féin, ní comhartha iontaofa iad na méadraigh chéanna ar a úsáidí agus atá an t-aschur sin i dtimpeallacht ghairmiúil iar-eagarthóireachta. Bíonn i gceist le meastóireacht dhaonna go n-iarrtar ar dhaoine a bhfuil saineolas acu ar an sprioctheanga aistriúchán a mheas chomh maith le tomhas a dhéanamh ar an méid saothair atá ag teastáil le haschur MA a cheartú. Ní mór a bheith in ann uirlisí teicneolaíochta aistriúcháin a mheas go hoibiachtúil de réir an dua iar-eagarthóireachta a bhaineann leo ionas gur féidir cinneadh stuama a dhéanamh i dtaca leis na huirlisí aistriúcháin a úsáidfear. Cinntíonn sé sin freisin gurb éard a bheidh mar thoradh ar na huirlisí aistriúcháin ná aistriúcháin ar lú an saothar a bhaineann le hiar-eagarthóireacht a dhéanamh orthu ná a mhacasamhail a bheadh i gceist dá n-aistreofaí na téacsanna céanna ó bhonn [22]. Caithfear caighdeán aistriúcháin na gcóras MA a thomhas, agus lena chois sin caithfear tomhas a dhéanamh ar a éifeachtaí nó ar a úsáidí agus atá córais MA d'aistritheoirí i dtimpeallachtaí sainiúla aistriúcháin chomh maith céanna [23].

Cé go ndearnadh roinnt taighde ar an meastóireacht dhaonna ar mheaisínaistriúchán na Gaeilge go dtí seo (ina gcuirtear innill OCBÁC, MAS agus MAN i gcomparáid le Google Translate), níl aon tuiscint shoiléir ann go fóill ar an gcaoi a dtéann MA i bhfeidhm ar obair an aistritheora Gaeilge. Is amhlaidh an scéal de bharr (a) a laghad úsáide a baineadh as MA Gaeilge i dtimpeallachtaí gairmiúla go dtí seo, agus (b) a laghad meastóirí atá ann faoi láthair a bhfuil an tacar riachtanach sainscileanna acu.

Áirítear le gníomhaíochtaí a bhaineann le meastóireacht dhaonna [24]:

i) i) Measúnú ar líofacht/ar shásúlacht gach abairte a aistríodh ar scála Likert
ii) ii) Measúnú rangúcháin ar cheann amháin acu seo a leanas:
– córais éagsúla MA (.i. saorchóras ar líne agus córas a cuireadh in oiriúint do réimse faoi leith) nó
– córas a cuireadh in oiriúint do réimse faoi leith agus atá á chur i bhfeidhm ar bhealaí difriúla bunaithe ar éagsúlachtaí i méideanna nó i gcineálacha na sonraí traenála
iii) iii) Saothar iar-eagarthóireachta a thomhas: dua ama (an t-am a thógtar le hiar-eagarthóireacht a dhéanamh) [25] dua teicniúil (líon na n-athruithe atá ag teastáil), agus dua cognaíoch (méid an tsaothair atá ag teastáil le hiar-eagarthóireacht a dhéanamh, arna thomhas ag rianaire súl) (26).

11.5.5 Oideachas agus Oiliúint

Tá roinnt réimsí inar gá infheistíocht a dhéanamh i sainoiliúint agus sainoideachas MA:

- Go hidéalach, ba chóir go mbeadh ag teicneolaithe meaisínaistriúcháin na scileanna iomlána seo a leanas: an innealtóireacht ríomhaireachta, an teoiric teangeolaíochta agus an Ghaeilge. Cinntíonn an meascán scileanna go bhfuil innealtóirí nó forbróirí córas MA sách cáilithe le hanailís a dhéanamh ar an aschur ó chóras meaisínaistriúcháin atá á fhorbairt faoi láthair, agus tiúnáil chuí a dhéanamh air dá réir.
- Ba chóir go mbeadh i ngach cúrsa tríú leibhéal i Léann an Aistriúcháin Gaeilge agus i gcúrsaí ábhartha araon modúil a chinnteoidh go mbeidh oiliúint leordhóthanach ar chéimithe ionas go mbeidh siad in ann úsáid a bhaint as teicneolaíochtaí reatha aistriúcháin (lena n-áirítear bogearraí cuimhne aistriúcháin agus uirlisí meaisínaistriúcháin). Áirítear leis sin oideachas i dtaca le bainistiú freagrach ar shonraí, le sonraí a chomhroinnt, agus leis an tábhacht a bhaineann le hathúsáid cuimhní aistriúcháin.
- Is scil shainiúil í iar-eagarthóireacht a dhéanamh ar aschur meaisínaistriúcháin Gaeilge nach mór do gach aistritheoir gairmiúil a bheith eolach uirthi [27].

Tagairtí

1. Dowling M, Cassidy L, Maguire E, Lynn T, Srivastava A, Judge J, et al. Machine Translation for Subtitling: A Large-Scale Evaluation. In: Pavlović Lažetić G, Krstev C, Obradović I, Duško V, editors. Proceedings of 9th International Conference on Language Resources and Evaluation (LREC), At Reykjavik, Iceland. Reykjavik, Iceland: European Language Resources Association (ELRA); 2014. p. 46–53.

2. Guerberof A. Productivity and quality in the post-editing of outputs from translation memories and machine translation. Int J Localis. 2008;7(1):11–21.

3. Walsh J. Enactments concerning the Irish language, 1922 - 2016. Dubl Univ Law J. 2016;39(2):449–66.

4. Dowling M, Cassidy L, Maguire E, Lynn T, Srivastava A, Judge J. Tapadóir: Developing a Statistical Machine Translation Engine and Associated Resources for Irish. In: Pavlović Lažetić G, Krstev C, Obradović I, Duško V, editors. The 4th LRL Workshop: "Language Technologies in support of Less-Resourced Languages." Poznan, Poland; 2015.

5. Forcada ML, Ginestí-Rosell M, Nordfalk J, O'Regan J, Ortiz-Rojas S, Pérez-Ortiz JA, et al. Apertium: A free/open-source platform for rule-based machine translation. Mach Transl. 2011;25(2):127–44.

6. Dowling M, Lynn T, Graham Y, Judge J. English to Irish Machine Translation with Automatic Post-Editing. In: Proceedings of the second Celtic Language Technology Workshop. 2016. p. 42--54.

7. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. Paper presented at the Advances in Neural Information Processing Systems (NIPS).

8. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. Paper presented at the 3rd International Conference on Learning Representations, Sab Diego.

9. Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. Paper presented at the Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas.

10. Dowling M, Lynn T, Poncelas A, Way A. SMT versus NMT: Preliminary comparisons for Irish. In: Proceedings of the AMTA 2018 Workshop on Technologies for MT of Low Resource Languages (LoResMT 2018). 2018. p. 12–20.

11.Arcan, M., Lane, C, Ó Droighneáin, E. & Buitelaar, P. (2016). IRIS: English-Irish machine translation system. Paper presented at the LREC 2016, Tenth International Conference on Language Resources and Evaluation, Portorož, Slovenia, 23-28 May.

12. Torregrosa, Pasricha, Masoud, Chakravarthi, Alonso, Casas, Arcan (2019) Leveraging Rule-Based MT Knowledge for Under-Resourced NMT Models. In: Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks https://www.mtsummit2019.com

13. Shaw, Stephen. 2013. Shallow transfer rule-based machine translation and the Irish-English language pair. Final Year Project Report, Trinity College Dublin

14. Dowling, Meghan. 2015. Rule Based Machine Translation for the Irish English pair using Apertium Final Year Project Report, Trinity College Dublin

15. O'Regan, Jim. 2017. English-Irish Machine Translation . M. Phil. Dissertation, Trinity College Dublin

16. Hamman, Ethan. 2019. Evaluating Google Translate for the English Irish language pair and Implementing Online Translation Checking. Final Year Project Report, Trinity College Dublin

17. Uí Dhonnchadha E. Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar. Dublin City University; 2009.

18. Eriguchi, A., Tsuruoka, Y., & Cho, K. (2017, jul). Learning to Parse and Translate Improves Neural Machine Translation. Paper presented at the Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Vancouver, Canada.

19. Bernardinello, G. (2019). Morphological Neural Pre- and Post-Processing for Slavic Languages. In: Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks https://www.mtsummit2019.com

20. Papineni K, Roukos S, Ward T, Zhu W-J. BLEU: a Method for Automatic Evaluation of Machine Translation. In: Isabelle P, Charniak E, Lin D, editors. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL). Philadelphia, USA; 2002. p. 311–8.

21. Denkowski, M., and A. Lavie. 2011. Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems. In Proceedings of the EMNLP 2011 Workshop on Statistical Machine Translation.

22. De Sousa SCM, Aziz W, Specia L. Assessing the post-editing effort for automatic and semi-automatic translations of DVD subtitles. In: Mitkov R, Angelova G, editors. International Conference Recent Advances in Natural Language Processing, RANLP. Hissar, Bulgaria; 2011. p. 97–103.

23. Fuji M, Hatanaka N, Ito E, Kamei S, Kumai H, Sukehiro T, et al. Evaluation Method for Determining Groups of Users Who Find MT "Useful." In: Proceedings of the Machine Translation Summit VIII (2001). 2001.

24. Moorkens J, Castilho S, Gaspari F, Doherty S. Translation Quality Assessment From Principles to Practice. Vol. 6, From Principles to Practice. 2018. 149–168 p.

25. Snover, M., B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul. 2006. A study of translation edit rate with targeted human annotation. In Proceedings of the Association for Machine Translation in the Americas, pages 223–231.

26. Stymne, Sara, Henrik Danielsson, Sofia Bremin, Hongzhan Hu, Johanna Karlsson, Anna Prytz Lil- Ikull, and Martin Wester. 2012. Eye Tracking as a Tool for Machine Translation Error Analysis. In Proceedings of the International Conference on Language Resources and Evaluation, Istanbul, Turkey.

27. Arenas, Ana (2008) Productivity and quality in the post-editing of outputs from translation memories and machine translation. Localisation Focus 7, pages 11-21

Caibidil 12
Aisghabháil Faisnéise

12.1 Céard is Aisghabháil Faisnéise ann?

Is ionann Aisghabháil Faisnéise (AF) agus modh le teacht ar fhaisnéis ó bhailiúchán mór foinsí sonraí. D'fhéadfadh sé gur doiciméid téacs, íomhánna, comhaid chlosamhairc a bheadh sna sonraí sin. Tá cineálacha éagsúla storála ann le haghaidh sonraí, lena n-áirítear eolairí cartlannaithe, stórtha logánta, bunachair sonraí inmheánacha, stórtha fearainn gréasáin agus mar sin de.

Mar réimse taighde, is gnách go ndírítear san AF ar fheabhas a chur ar chruinneas na dtorthaí ábhartha úsáideacha a bhíonn ar chuardach eochairfhocal. Is gnách go ndéantar torthaí a rangú de réir a ábhartha agus atá siad i dtaca leis an gcuardach. Tá cineálacha éagsúla feidhmchlár AF ann, agus bainimid tairbhe as roinnt cineálacha sainiúla inár saol laethúil. Cuir i gcás, ar inneall cuardaigh gréasáin, d'fhéadfadh sé go mbeadh de thoradh ar an gcuardach eochairfhocail "ranganna oíche i mBaile Átha Cliath" liosta leathanach ina dtugtar faisnéis i dtaca le ranganna do dhaoine fásta, ranganna oíche don phobal, cúrsaí oíche atá á reáchtáil ag ollscoil, srl. Dá ndéanfaí cuardach ar "tuarascálacha soláthróirí 2015" i mbunachar sonraí gnó, bheifí ag dréim leis go ndéanfaí aisghabháil ar gach doiciméad maidir le tuarascálacha a bhaineann le soláthróirí don bhliain 2015.

Mar sin féin, mar thoradh ar an méadú ollmhór ar an méid téacs ar líne a bhfuil fáil air agus ar an éileamh ar rochtain ar chineálacha éagsúla faisnéise, tá breis réimsí a bhaineann le AF ann ná díreach an aisghabháil shimplí faisnéise. Ar chuid de na samplaí tá freagairt ceisteanna, aimsiú agus rianú topaicí, achoimriú, aisghabháil ilmheán (.i. íomhánna, físeáin agus ceol), innealtóireacht bogearraí, faisnéisíocht cheimice agus bhitheolaíochta, struchtúrú téacs, mianadóireacht téacs, agus géanómaíocht (1).

12.2 Cén tábhacht atá le hAisghabháil Faisnéise?

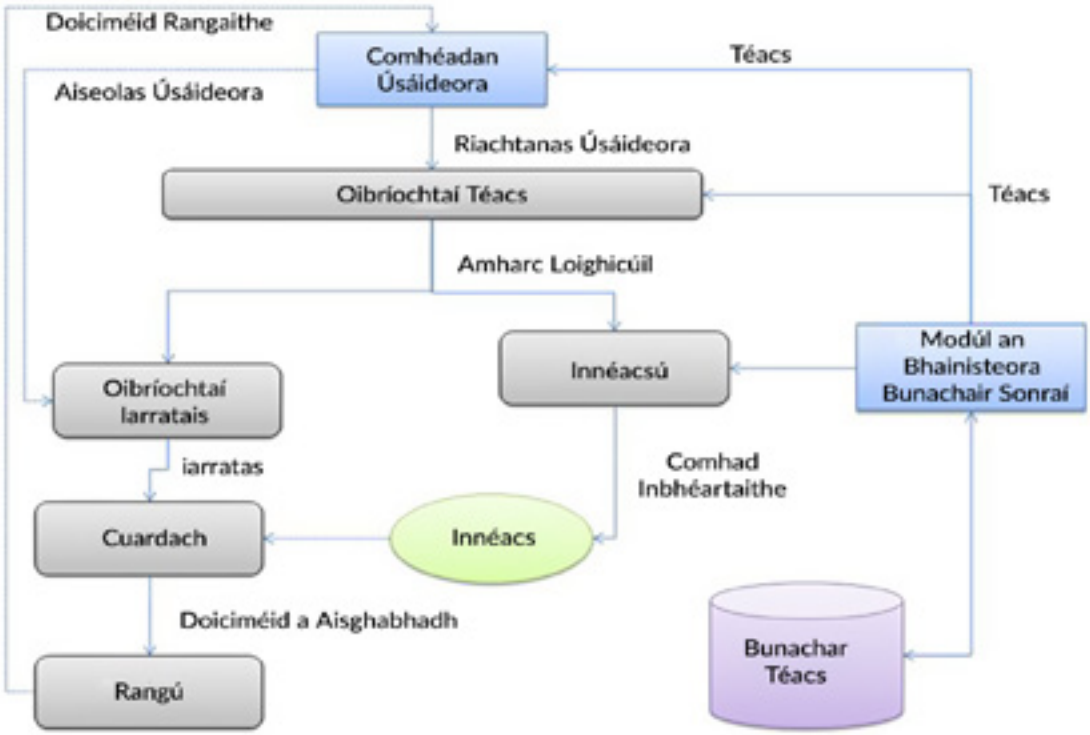
Faoi mar atá an méid sonraí atá ar fáil ar líne agus i cartlanna príobháideacha ag dul i méid, is amhlaidh atá an gá leis an teicneolaíocht chuardaigh chruinn ag éirí níos follasaí le caoi a thabhairt d'úsáideoirí teacht ar an bhfaisnéis atá uathu. Níl an gá le córais iontaofa AF teoranta le cuardach ginearálta de chuid an gréasáin dhomhanda, áfach. Láithreáin ghréasáin aonair, ina bhfuil líon mór leathanach gréasáin nó líon mór sonraí cartlannaithe, bíonn innill chuardaigh chruinne de dhíth orthu.

Ina theannta sin, faoi mar a théann an bunachar doiciméad atá ag gnólacht nó ag eagraíocht i méid, is amhlaidh a théann i méid freisin an gá le huirlisí leis na sonraí sin a brabhsáil go furasta chomh maith. Bíonn cineálacha éagsúla AF iontaofa ag teastáil de réir na gcineálacha inneachair a bhíonn á gcoinneáil nó á gcartlannú ag eagraíocht, agus de réir chuspóir na n-iarrachtaí aisghabhála. Cuir i gcás, b'fhéidir go mbeadh ar ghnólacht taifid custaiméirí a chuardach go cruinn d'fhonn tuarascálacha a chruthú nó anailís a dhéanamh ar thuarascálacha. Bíonn taifid de leabhair, foilseacháin, irisí, nuachtáin, srl. ag leabharlann. Bíonn cartlann d'agallaimh, cláir faisnéise, tuairiscí, comhaid fotheideal, agus mar sin de ag craoltóir.

I gcomhthéacs na Gaeilge, tá cásanna go leor ann ina bhfuil gá le córais fheabhsaithe AF. Is féidir le córais AF a bhfuil tiúnáil déanta orthu cuidiú le bainistíocht sonraí níos éifeachtaí a bhaint amach d'eagraíochtaí gnó, taighde ar líne agus taighde inmheánach rialtais, taighde acadúil ar líne agus taighde cartlainne, aisghabháil cartlainne ag institiúidí meán ar inneachar closamhairc, mar aon le taighde ar líne níos leithne ar chomhaid ilmheán.

12.3 Conas a oibríonn sé?

Bíonn Aisghabháil Faisnéise (AF)/Inneall Cuardaigh bunúsach sách neamhspleách ar theanga, agus ní dhírítear iontu ach ar innéacsú agus rangú leathanach agus doiciméad de réir mhinicíocht choibhneasta na dtéarmaí a úsáidtear sa chuardach eochairfhocail sa bhailiúchán doiciméad atáthar a chuardach. Mar sin féin, bíonn uirlisí iontaofa cuardaigh spleách ar theanga. Is é sin le rá, deartar iad ionas go gcuirtear san áireamh faisnéis theangach na teanga atá sna doiciméid nó sna leathanaigh atá á gcuardach. Is éard is cúis le rathúlacht an taighde reatha AF ná an comhtháthú ar mhodhanna de chuid na próiseála teanga nádúrtha a thugann caoi don chóras a bheith níos feasaí agus níos tuisceanaí ar theanga. Bíonn na modhanna sin ina gcuidiú le feabhas a chur ar ionadaíocht chomh maith le feabhas a chur ar an gcaoi a dtuigtear na hiarratais agus na doiciméid ionas go dtiocfaidh feabhas ar an meaitseáil.



Fíor 1. Forléargas Samplach ar Phróiseas Aisghabhála Faisnéise (Buscaldi, 2011)

Tá innill chuardaigh ardfheidhmiúla ann le haghaidh teangacha a bhfuil acmhainní go leor curtha ar fáil dóibh amhail an Béarla, an Ghearmáinis, an Fhraincis. Tá feabhas suntasach tagtha ar na hinnill chuardaigh sin le blianta beaga anuas ó thaobh chéatadán na leathanach gréasáin nó na ndoiciméad ábhartha a aisghabhtar trí bhíthin cuardach eochairfhocail. Tá an dul chun cinn sin ar an teicneolaíocht AF teoranta le teangacha a bhfuil acmhainní go leor curtha ar fáil dóibh, áfach. Is annamh a bhíonn fáil ar an teicneolaíocht AF agus í curtha in oiriúint do theangacha atá ar bheagán acmhainní ar nós na Gaeilge. Cé go bhféadfadh sé go gcuirfí comhéadan Gaeilge ar fáil in innill chuardaigh ghréasáin ar nós Google, níl le sonrú ar a gcuid uirlisí cuardaigh siúd aon bhuntuiscint dhomhain ar an teanga chun go mbeidh caighdeán na dtorthaí ag teacht le macasamhail na dteangacha eile a dtugtar tacaíocht níos fearr dóibh.

12.4 A bhfuil déanta i leith na Gaeilge go dtí seo

Rinneadh roinnt iarrachtaí teoranta logánta córais AF a chruthú le haghaidh láithreán gréasáin beag (.i. www.gaeilge.ie de chuid Fhoras na Gaeilge). Go dtí seo, áfach, ní dhearnadh aon taighde cuimsitheach a chur i gcrích nó a thuairisciú i dtaca le modhanna AF reatha úrscothacha le haghaidh na Gaeilge a chur i bhfeidhm ná a mheasúnú, ná i dtaca le forbairt modhanna nua AF atá curtha in oiriúint don Ghaeilge ach an oiread. Sa chaibidil seo, mar sin, caitear léas ar na réimsí uile den taighde agus den fhorbairt ar an aisghabháil faisnéise ar gá aird a dhíriú orthu amach anseo.

Is tábhachtach cuimhneamh air go bhfuil cainteoirí Gaeilge ag úsáid uirlisí cuardaigh ar líne agus as líne a ndearnadh tiúnáil agus optamú orthu chun inneachar digiteach Béarla a chuardach. Mar sin de, tá teorainn le cumas na n-uirlisí sin téacs nó comhaid fuaime Ghaeilge a phróiseáil i gceart, agus bíonn torthaí míchruinne agus neamhiomlána mar thoradh orthu ar an mórgóir dá bhrí sin.

12.5 Moltaí

Beidh an taighde agus an fhorbairt ar chórais AF Ghaeilge ina ndul chun cinn suntasach i dtaca le bainistíocht a dhéanamh agus rochtain a fháil ar acmhainní Gaeilge a dhéanfaidh leas bhaill an phobail, leas corparáidí teilifíse agus raidió, leas gnólachtaí, agus leas rialtas. Siúd a leanas na príomhréimsí a mholtar mar spriocanna tosaíochta i gcaitheamh na 5-10 mbliana atá le teacht.

12.5.1 Cuardach agus Aisghabháil Téacs atá tiúnáilte le haghaidh na Gaeilge

Is gá córais AF ar ardchaighdeán a thiúnáil do theanga

shainiúil an iarratais. Cíortar go leor de na bunacmhainní teicneolaíochta teanga atá ag teastáil le haghaidh oiriúnú bunúsach a dhéanamh ar AF Gaeilge i gcaibidlí roimhe seo. Caithimid léas anseo ar chuid de na réimsí nach mór iad a chur i bhfeiliúint do thaighde sainiúil AF. Is fiú a lua, áfach, go mbíonn innill rathúla IR ag brath ar a lán nithe eile le cois chomhtháthú na n-uirlisí a áirítear thíos:

- Úsáidtear uirlisí amhail tamhantóirí nó leimeoirí le teacht ar thamhan eochairfhocail agus le cuidiú le hinneall AF aisghabháil a dhéanamh ar dhoiciméid ina bhfuil focail eile ag a bhfuil baint mhoirfeolaíoch leis an tamhan céanna, .i. seachas an cuardach eochairfhocail a theorannú leis an bhfoirm infhillte amháin. Cuir i gcás, i measc na dtorthaí ar chuardach ar an eochairfhocal teicneolaíocht ba chóir go mbeadh leathanaigh nach bhfuil ach na foirmeacha infhillte teicneolaíochta nó teicneolaíochtaí le sonrú iontu chomh maith. (Féach Caibidil 5 le tuilleadh tagairtí a fháil).
- A dhála sin, is féidir faisnéis shéimeantach (cialla focal) a bhaint as líonraí léacsacha séimeantacha amhail WordNet chun inneall cuardaigh a chur ar an eolas faoi ghaolta idir focail, amhail comhchiallachas [2]. Cuir i gcás, ar an toradh is airde ar chuardach ar an eochairfhocail Béarla 'Irish language technology' tá leathanach gréasáin ina ndéantar cur síos ar 'app' nua foghlama teanga. Cé nach bhfuil an focal 'technology' le sonrú sa doiciméad ardrangaithe a thagann chun cinn, tá an córas AF Béarla sách éirimiúil chun an gaol séimeantach idir 'app' agus 'technology' a aithint. D'fhéadfaí an-leas a bhaint as Líonra Séimeantach na Gaeilge (féach Caibidil 4) i gcomhair sonraí den sórt seo le haghaidh córais AF don Ghaeilge.
- Ina cheann sin, úsáidtear parsálaithe ar uairibh le cuidiú le modhanna AF, trí chiseal breise faisnéise i dtaca le struchtúr an téacs sna doiciméid atá á gcuardach a chur ar fáil, nó leis na gnéithe tábhachtacha de chuardach atá bunaithe ar cheist a bhaint amach [3]. Bheadh na parsálaithe a cíoradh i gCaibidil 5 úsáideach le struchtúr téacs a aithint agus a scagadh chun feabhas a chur ar innill chuardaigh optamaithe don Ghaeilge. Bhainfeadh an cur chuige seo le hiarratais atá bunaithe ar urlabhra chomh maith.

Aon taighde a dhéantar ar an aisghabháil téacs ar dtús, is féidir síneadh a chur leis sin trí chuardach a dhéanamh ar chartlanna ilmheán a cruthaíodh de láimh nó a bhfuil tras-scríbhinní a baineadh amach go huathoibríoch ceangailte leo. Bheadh an cineál seo teicneolaíochta thar a bheith luachmhar d'eagraíochtaí amhail stáisiúin teilifíse, stáisiúin raidió, dreamanna a bhíonn i mbun taighde agus cartlannú daonnachtaí digiteacha de chuid na Gaeilge, agus mar sin de. I gcásanna áirithe, d'fhéadfadh sé nár ghá córais a chruthú ó bhonn ach ina ionad sin d'fhéadfaí creataí foinse oscailte atá ann cheana féin amhail Terrier a chur in oiriúint dá ndéanfaí uirlisí próiseála teanga nádúrtha, atá sainiúil don Ghaeilge, a chuimsiú iontu go cuí.

12.5.2 Achoimriú na dTorthaí Aisghabhála Faisnéise

Nuair a chuirtear torthaí (.i. doiciméid) faoi bhráid úsáideora, bíonn cuidiú ag teastáil ón úsáideoir i dtaca leis na doiciméid is ábhartha lena c(h)uardach a dhéanamh amach, gan gach doiciméad a scrúdú ceann ar cheann. Réiteach a sholáthraíonn innill chuardaigh go minc is ea blúirí a úsáid, ina mbíonn teideal doiciméid agus achoimre 2 nó 3 líne ar an doiciméad de ghnáth. Tríd is tríd, is éard a bhíonn i gceist leis na teicnící ná go ndírítear ar rogha abairtí de chuid an doiciméid a aithint, abairtí a mheastar a bheith ionadaíoch ar an inneachar. Bíonn an tasc achoimriúcháin sin ina thasc doiligh san fhorbairt AF uaireanta mar go mbíonn sé ag brath ar theicnící cliste de chuid na Próiseála Teanga Nádúrtha (féach Caibidil 5) agus de chuid na Giniúna Teanga Nádúrtha (féach Caibidil 6) a bhaineann go sainiúil leis an teanga atá i gceist, amhail samhaltú, giniúint, nó parsáil teanga.

Bheadh achoimriú na ndoiciméad Gaeilge a aimsítear i dtorthaí cuardaigh ina chuidiú luachmhar le háiseanna cuardaigh inneachair ar líne a fheabhsú, agus bheadh sé ina chuidiú suntasach freisin le feabhas a chur ar inneachar oidhreachta cartlannaithe inmheánach a mbíonn cuma rómhór dholáimhsithe air go minic lena phróiseáil go furasta.

12.5.3 Aisghabháil Faisnéise Thrasteangach

Is réimse taighde é an AF trasteangach ina gcuirtear le chéile teicnící de chuid an mheaisínaistriúcháin (MA) agus teicnící de chuid na haisghabhála faisnéise. Faoi láthair, táthar ag díriú go mór sa réimse seo ar réiteach a fháil ar an deacracht a bhaineann lena laghad sonraí digiteacha atá ar fáil le haghaidh teangacha atá ar bheagán acmhainní, trí thorthaí breise ábhartha a aimsiú ar chuardach eochairfhocail i dteangacha eile a bhfuil breis acmhainní ar fáil dóibh.

Cuir i gcás, tugann an córas AF Bascaise 'Zientzia' [4–7] caoi don úsáideoir cartlann d'ailt eolaíochta sa Bhascais a chuardach trí bhíthin eochairfhocail Bhaiscise a chur isteach. Déantar an chéad tacar de na doiciméid Bhascaise ar a bhfuil an rangú is airde a chur faoi bhráid an úsáideora. Ina theannta sin, bíonn torthaí i mBéarla agus i Spáinnis ann chomh maith, a bhuíochas sin ar an gcomhthathú ar chórais MA Bascais-Béarla agus Bascais-Spáinnis. De bharr an chur chuige sin, leathnaítear an

inrochtaineacht ar shonraí atá ar fáil don úsáideoir a bheartaíonn ar chuardach a dhéanamh sa Bhascais amháin. Déanann an córas innéacsú ar láithreáin gréasáin eolaíochta eile agus cuireann sé ar fáil mar roghanna de thorthaí tánaisteacha iad. A dhála sin, is córas AF trasteangach greásánbhunaithe é an t-inneall cuardaigh ar líne 'Elebila™' a thugann caoi don úsáideoir cuardach a dhéanamh i rogha de cheithre theanga (Bascais, Béarla, Spáinnis, Fraincis) agus gach toradh ábhartha a aisghabháil. Tá an rogha ann comhchiallaigh agus foirmeacha infhillte comhchiallach a chur san áireamh d'fhonn síneadh a chur leis an gcuardach. In Elebila, cuirtear ar fáil d'úsáideoirí breiseán le haghaidh Word nó Open Office mar aon le breiseán leathnaigh gréasáin.

De bhrí go bhfuil an méid sonraí Gaeilge atá ar fáil ar líne sách teoranta go fóill, moltar go láidir go ndéanfar taighde ar chóras trasteangach AF a fhorbairt amhail na córais siúd atá ar fáil do chainteoirí Bascaise. Dá bhforbrófaí a leithéid de chóras ina gcuimseofaí lena chéile na teicneolaíochtaí meaisínaistriúcháin Gaeilge-Béarla atá ann cheana agus a bheidh ann amach anseo (féach Caibidil 11), bheadh an t-úsáideoir in ann eochairfhocail a chuardach i nGaeilge, torthaí ardrangaithe le haghaidh doiciméid Ghaeilge ar líne a aisghabháil, agus doiciméid Bhéarla ina bhfuil faisnéis ábhartha a d'fhéadfadh a bheith úsáideach don úsáideoir a aisghabháil chomh maith céanna.

12.5.4 Aisghabháil Inneachair Labhartha Gaeilge

A thúisce a chuirfear ar fáil le haghaidh na Gaeilge córas aitheanta cainte uathoibríche (ACU) (féach Caibidil 9) atá iontaofa láidir, beifear in ann na tras-scríbhinní ACU a innéacsú lena n-úsáid i gcóras AF, agus beifear in ann cartlanna fuaime agus físeán a chuardach dá réir.

Ar dtús, is féidir cuimhneamh ar na tras-scríbhinní mar leagan malartach d'ionchur téacsúil don chóras AF, agus is féidir an t-oiriúnú tosaigh le haghaidh na Gaeilge a chur i bhfeidhm láithreach bonn (trí úsáid a bhaint as tamhantóirí agus liostaí stadfhocal [8]. Is féidir tuilleadh córais AF thrasteangacha a thástáil d'fhonn réimse an chuardaigh a leathnú le go n-áireofaí cartlanna ilmheán i dteangacha eile ann. Tras-scríbhinní ACU atá ar chaighdeán aitheanta 100%, bíonn siad éagsúil fós féin le téacs traidisiúnta, arae bíonn abairtí níos giorra á n-úsáid, cuirtear isteach ar an gcaint, bíonn neamhlíofachtaí ann, agus mar sin de, agus daoine i mbun cainte.

Dá bhrí sin, chun Aisghabháil Inneachair Labhartha a thabhairt chun cinn ó thaobh forbairt córais agus feabhas caighdeáin de, beidh gá le hiniúchadh ar na difríochtaí idir an teanga labhartha agus an teanga scríofa, agus seans go mbeidh gá le tiúnáil shainiúil a dhéanamh ar uirlisí freisin. Cuir i gcás, seans go mbeadh an pharsáil abairte ar inneachar labhartha ní ba chosúla leis an bparsáil ar inneachar meán sóisialta nach ndearnadh eagarthóireacht air le hais mar a bheadh an pharsáil ar ghnáth-théacs nuachtáin, de bharr neamhlíofachtaí agus frásaí neamhiomlána a bheith san inneachar labhartha agus san inneachar meán sóisialta faoi seach.

References

1. Allan J, Aslam J, Belkin N, Buckley C, Callan J, Croft B, et al. Challenges in Information Retrieval and Language Modeling. Massachusetts; 2002.

2. Rosso P, Ferretti E, Jiménez D, Vidal V. Text Categorization and Information Retrieval Using WordNet Senses. In: Proceedings of the Second Global Wordnet Conference. Brno; 2004. p. 299–304.

3. Maxwell TK, Oberlander J, Croft BW. Feature-Based Selection of Dependency Paths in Ad Hoc Information Retrieval. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Sofia: Association for Computational Linguistics; 2013. p. 507–16.

4. Saralegi X, López de Lacalle M. Estimating Translation Probabilities from the Web for Structured Queries on CLIR. In: Proceedings of Advances in Information Retrieval, 32nd European Conference on (IR) Research, (ECIR) 2010, Milton Keynes, UK, March 28-31, 2010 Proceedings. Milton Keynes; 2010. p. 586–9.

5. Saralegi X, López de Lacalle M. Dictionary and monolingual corpus-based query translation for basque-english CLIR. In: Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010. Valetta; 2010. p. 1353–8.

6. Saralegi X, López de Lacalle M. Comparing different approaches to treat translation ambiguity in CLIR: structured queries vs. target co-occurrence based selection. In: Proceedings of International Workshop on Database and Expert Systems Applications, DEXA. Linz; 2009. p. 398–404.

7. Saralegi X, Alegria I. Similitud entre documentos multilingües de carácter científico-técnico en un entorno Web. Proces del Leng Nat. 2007;39:71–8.

8. Uí Dhonnchadha E, van Genabith J. Scaling an Irish FST morphology engine for use on unrestricted text. In: Yli-Jyrä A, Karttunen L, Karhumäki J, editors. Finite-State Methods in Natural Language Processing 5th International Workshop, FSMNLP 2005, Helsinki, Finland, 2005 Revised Papers. Berlin: Springer International Publishing; 2006. p. 247--258.

Caibidil 13
Áiseanna don Phobal Mór

13.1 Cad iad féin?

Tá bunathruithe ag teacht ar stair an chine dhaonna. Tá teicneolaíocht na hurlabhra agus na teanga fite fuaite chomh dlúth sin sa ghnáthshaol gur ar éigean a thugaimid faoi deara iad. Mar a deirtear sa Réamhrá, is cuid dár saol iad na córais dhigiteacha nuair a thagann sé go cúrsaí oibre agus eile – spraoi, cumarsáid, eolas a lorg, foghlaim, plé le seirbhísí agus an timpeallacht ina measc.

Is é an aidhm atá ag an bPlean Digiteach ná áiseanna a fhorbairt chun go mbeidh an Ghaeilge ar láimh sa ghnáthshaol – saol atá de shíor ag dul i muinín córas éagsúla digiteacha. Is den riachtanas é go mbeidh an Ghaeilge sna réimsí saoil ar fad, m.sh. réimse na trádála, an dlí, caitheamh aimsire agus eile. Má ghearrtar siar ar a réimsí úsáide leáfar stádas na teanga dá réir. Má bhíonn lúb ar lár sa teicneolaíocht beidh ar dhaoine dul níos mó i muinín an Bhéarla ina saol laethúil, rud a chuirfidh go mór le cúngú na teanga. Mar mhalairt de sin thabharfadh teicneolaíocht shofaisticiúil urlabhra agus teanga deiseanna úrnua chun an Ghaeilge a neadú i ngach réimse saoil.

Beidh éagsúlacht mhór sna córais dhigiteacha a bheidh ar fáil don phobal agus tiocfaidh fás orthu sin de réir mar a théann an teicneolaíocht chun cinn. Ina measc siúd beidh:

- feidhmchláir neamhspleácha ina mbaintear feidhm as modúil teangeolaíochta, m.sh., seiceálaí gramadaí chun dul le próiseálaí focal;
- córais chasta a bhraitheann ar bhuntheicneolaíochtaí, m.sh., córas deachtúcháin a thagann i dtír ar aithint cainte;
- comhchórais chasta, m.sh., aistriúchán ó theanga labhartha go teanga labhartha – cónascadh d'aithint cainte, ríomhaistriúchán agus sintéis chainte sa dara teanga. Bheadh fotheidealú fíor-ama mar chuid de seo ionas go bhféadfadh duine éisteacht le píosa cainte agus í a léamh as Gaeilge fad is atá sí á cur i láthair i dteanga eile;
- córais idirghníomhacha fhíorúla atá leabaithe laistigh de ghníomhaí corpraithe mar róbat lena ndéantar idirphlé. Cuirtear leis an gcumarsáid trí léiriú láimhe agus éadain.

Níl teorainn leis na féidearthachtaí atá ann chun teicneolaíocht na hurlabhara agus na teanga a lonnú i measc an phobail agus ní dhéanfar sa chaibidil seo ach blaiseadh a thabhairt orthu. Ós rud é go bhfuil dhá réimse atá ríthábhachtach, oideachas agus rochtain dóibh siúd atá faoi mhíchumas, pléite i gCaibidlí 14 agus 15, díreofar anseo ar réimsí eile.

13.2 Cén tábhacht atá leo agus cé dó iad?

Tá forbairt na n-áiseanna seo tábhachtach do phobal uile na Gaeilge, dóibh siúd a bhfuil an Ghaeilge lárnach ina gcuid oibre, d'fhoghlaimeoirí agus do dhaoine a bhfuil spéis acu sa Ghaeilge ar chúis amháin nó ar chúis eile. Tá plé anseo ar fheidhmchláir todhchaíocha mar aon le teicneolaíochtaí seanbhunaithe a rachaidh i bhfeidhm ar ghnéithe éagsúla dár saol.

13.2.1 Imeachtaí an Lae

Glactar go forleathan leis an teicneolaíocht ar mhaithe le luas, éifeachtúlacht agus áisiúlacht agus is comhpháirt lárnach é an nascadh cruthaitheach le hurlabhra agus teanga.

Sa bhaile: timpeallacht chliste. Tá ár dtithe ag éirí níos 'cliste' in aghaidh an lae. Tá ag méadú ar ár gcumas fearais agus cúraimí baile a riar le teicneolaíochtaí guthbhunaithe, m.sh. cláir raidió nó theilifíse a aimsiú, soilse a rialú, córais slándála, teasa nó siamsaíochta a riar. Is minic gur cianrialú le fón póca a bhíonn i gceist. Is rochtain thábhachtach é a leithéid fotheidealú uathoibríoch dóibh siúd a bhfuil máchail éisteachta orthu. Éascaíonn sé iltascáil sa mhéad is gur féidir linn cumarsáid a dhéanamh ó bhéal a fhad is atá ár lámha nó ár súile gafa le rud éigin eile. Cheana féin, mar shampla, labhraíonn daoine le córais suite domhanda (GPS) a fhad is atá siad ag tiomáint ina gcarranna nó lorgaíonn siad comhairle ó bhéal ó chúntóirí cainte agus iad i mbun cócaireachta.

San ionad oibre: cúntóirí pearsanta. Baintear úsáid go forleathan as modúil phróiseála teanga nádúrtha chun ábhar a chumadh, litriú agus gramadach a sheiceáil, profaí a léamh agus eagarthóireacht a dhéanamh. Ní mór iad seo a leasú agus a mhéadú (tuilleadh in §13.4).

De réir mar a thagann bunteicneolaíochtaí nua chun

cinn cuirfear le cumadóireacht ábhair, m.sh. téacsanna a dheachtú, éisteacht profaí a dhéanamh chun cur leis an éifeacht a théann le léamh profaí, agus i bhfad na haimsire, feidhmchláir chun achoimriú a dhéanamh ar théacs – rud a chuirfidh go mór le bearta móra faisnéise a phróiseáil.

Le comhchórais chasta tagann fás leanúnach ar líon na bhfeidhmchlár úsáideach. Sample de seo is ea nuair a chuirtear teicneolaíocht an aistriúcháin sa chomhchóras is féidir faisnéis a bhailiú trí Ghaeilge, faisnéis a bhfuil a bunleagan stóráilte i dteanga éigin eile. Le teicneolaíocht na hurlabhra d'fhéadfaí an fhaisnéis seo a lorg agus a fháil ó bhéal.

Imeachtaí sóisialta agus siamsaíochta. Trí úsáid a bhaint as gréasáin shóisialta ar fhearais shoghluaiste is féidir linn a bheith i dteagmháil lenár gcairde, cruthaímid gréasáin nua shóisialta, coimeádtar ar an eolas sinn faoi imeachtaí atá ag tarlú, faoina bhfuil ar siúl i measc ár gcairde agus ceadaítear dúinn cairde nua a lorg. Úsáidtear gníomhairí cliste chun eagar a choimeád ar ár ndialann shóisialta, m.sh., ticéid nó eile a chur in áirithe.

In imeachtaí eile cosúil le cuairt a thabhairt ar mhúsaem nó gailearaí ealaíne beidh teacht ar níos mó idirghníomhaíochta le taispeántáin, m.sh., d'fhéadfadh gníomhaire a bheith leabaithe i gcarachtair stairiúla a bheadh in ann labhairt amach agus ceisteanna a fhreagairt (féach Caibidil 10). Tá na leaganacha atá anois ann réasúnta simplí ach le forbairt na Réaltachta Breisithe beifear in ann scéalta ár muintire a chur ina steillbheatha. Samhlaigh thú féin ag siúl thart ar an mBlascaoid agus an t-oileán a bhlaiseadh faoi mar a bhí sí agus í i lár a maitheasa!

Tá cluichí idirghníomhacha ilmhódacha tagtha go mór chun cinn mar chaitheamh aimsire do gach aoisghrúpa agus tugann siad deis do dhaoine spraoi a bheith acu as Gaeilge. Faoi láthair baintear úsáid i gcluichí móréilimh as cúntóirí cainte i gcomhair cluichí teaghlaigh agus cluichí do pháistí, m.sh. tráth na gceist nó cathaoireacha ceoil. Ós rud é go bhfuil saol digiteach ón gcliabhán ag páistí an lae inniu agus gan cur amach acu ar aon saol eile, beidh siad ag súil le bréagáin 'chliste' a bheidh ábalta éisteacht leo agus iad a fhreagairt. De réir mar a fhásfaidh siad aníos ba cheart go mbeadh fáil acu ar chláir mar 'CoderDojo ' agus 'Hour of Code ' as Gaeilge chun go spreagfaí iad chun spraoi a bhaint as cluichí agus ábhar dá gcuid féin a fhorbairt as Gaeilge.

Plé leis an Stát agus le Comhlachtaí Poiblí. De thairbhe na teicneolaíochta urlabhra agus teanga cuirtear ar chumas an Stáit agus na gcomhlachtaí poiblí a ndualgais faoi Acht na dTeangacha Oifigiúla a chomhlíonadh agus seirbhísí trí Ghaeilge a chur ar fáil don phobal uile. Teicneolaíocht ríthabhachtach is ea an ríomhaistriúchán lena gcuirfí na doiciméid uile ar fáil go dátheangach. Is bunáis í teicneolaíocht na hurlabhra agus na teanga trína gcuirfí faisnéis, foirmeacha agus eile ar fáil go saoráideach agus a thabharfadh an deis tascanna a chomhlíonadh ó bhéal nó i bhfoirm téacs.
Tá sé de dhualgas ar an Stát deimhin a dhéanamh de go bhfuil teacht go saoráideach ag daoine ar eolas go háirithe iad siúd a bhfuil míchumas orthu (féach Caibidil 15). I measc na dteicneolaíochtaí atá ríthábhachtach tá sintéis téacs-go-hurlabhra dóibh siúd a bhfuil máchail súl nó máchail chainte orthu; urlabhra-go-téacs chun fotheidealú a dhéanamh ar chaint dóibh siúd atá bodhar; agus gléasanna achoimre téacs mar áis dóibh siúd a bhfuil easnamh airde orthu, m.sh. chun dul trí doiciméid chasta.

Sláinte. Tá bunathruithe ag teacht ar sholáthairtí sláinte. Tá méadú ag teacht ar chórais ina n-úsáidtear cúntóirí cainte chun ceisteanna faoi chúrsaí sláinte a fhreagairt. Is minic a úsáidtear róbait i gcomhair seandaoine nó daoine atá faoi éagumas chun a gcuid neamhspleáchais a chaomhnú nó chun comhluadar a choimeád leo nó chun mearbhú dóibh faoi chógais, coinní, srl. Is féidir leis na róbait seo, a rialaítear trí orduithe gutha, miontascanna a dhéanamh do dhaoine a bhfuil m.sh. galar néarón luadrach orthu, leithéid an citeal a chur síos agus cupán tae a dhéanamh, mar shampla. Chabhródh sé seo go mór lena neamhspleáchas.

Táthar ag baint níos mó úsáide anois as córais (aipeanna fóin, mar shampla) a dhéanann monatóireacht ar chaint an othair chun súil a choimeád ar a shláinte fhisiceach nó ar a mheabharshláinte. Féachtar ar phrosóid na cainte, mar shampla, chun comharthaí dúlagair a aithint go luath (féach Caibidil 2.4.3), rud a thugann seans do theiripeoir idirghabháil a dhéanamh in am is i dtráth.

13.2.2 Pobail a Thógáil aguss a Nascadh: Teicneolaíocht ar son na Maitheasa

Cruthú Pobail. Cruthaítear pobail nua trí ghreasáin dhigiteacha nó ardáin meáin shóisialta a úsáideann daoine as a stuaim féin, rud a thugann daoine le chéile chun ábhair spéise a roinnt – ceol nó spórt, mar shampla. Is as Béarla den chuid is mó a bhíonn na hardáin sin cé go bhfuil Gaeilge á cur ar chuid acu faoi láthair (féach 13.4.3 thíos). Cuirfidh an chumhacht bhreise a bhaineann le huirlisí téacs agus urlabhra le cumas sainghrúpaí an Ghaeilge a úsáid mar mheán cumarsáide chun díriú ar ábhair eile seachas ar an nGaeilge í féin. Sa ré atá caite bhíodh na gréasáin seo logánta ach leis an nglúin nua atá eolach ar an teicneolaíocht leathnaítear na gréasáin seo ionas go mbíonn daoine ó na ceithre hairde mar bhaill iontu agus iad á n-ionramháil de réir a mianta féin.

Gréasáin Phobail. Faoi mar a pléadh i gCaibidil 1 tá pobal na Gaeilge scaipthe i gceantair bheaga Ghaeltachta atá scartha óna chéile agus tá roinnt daoine eile ina mbaill de ghrúpaí

beaga Gaelgeoirí atá lonnaithe ar fud na tíre agus thar lear. Tá gréasán tábhachtach idir na grúpaí seo cruthaithe ag TG4 agus ag Raidió na Gaeltachta agus ceangal déanta ag na meáin seo leis an bpobal mór. Tá teicneolaíocht na hurlabhra agus na teanga ina lóchrann a threoróidh na pobail chun sainardáin éagsúla a úsáid agus a cheapadh, m.sh., ardáin ina bhfuil gníomhairí éagsúla corpraithe le Réaltacht Bhreisithe. Bheadh sé d'acmhainn ag a leithéid seo de theicneolaíocht na gréasáin phobail a mhéadú agus a neartú agus féiniúlacht a thabhairt dóibh le go bhfeicfeadh daoine iad féin mar chuid de 'phobal na Gaeilge' a bhfuil oidhreacht shaibhir acu le roinnt.

13.3 Conas a oibríonn sé?

Mar a pléadh thuas, beidh gá le feidhmchláir ilchineálacha a bheadh freagrúil do riachtanais éagsúla. Tugann an Plean Digiteach seo an deis, ní hamháin chun bheith ar aon chéim le teicneolaíochtaí atá ann cheana féin don Bhéarla, ach chun an chruthaitheacht a spreagadh chun feidhmchláir ar leith a fhorbairt a bheidh freagrúil do shainriachtanais na Gaeilge sa 21ú céad.

13.4 A bhfuil déanta i leith na Gaeilge go dtí seo

Tá roinnt feidhmchlár ar fáil cheana féin don phobal agus tá roinnt eile á bhforbairt. I measc na samplaí thíos tá uirlisí téacsbhunaithe, m.sh., uirlisí lena n-éascaítear léamh profaí, uirlisí le seiceáil ar chruinneas agus ar inléiteacht, tuarthéacs, agus cuardach foclóra. I measc na n-uirlisí urlabhartha tá leathanach gréasáin, léitheoir gréasáin agus aip Android a léann amach an téacs ar líne. Réimse eile atá tábhachtach ná logánú, rud a chuireann ardáin atá bunaithe cheana féin in oiriúint do chomhthéacs na Gaeilge.

13.4.1 Uirlisí téacsbhunaithe

Uirlisí profála téacs. Ina measc siúd tá seiceálaithe litrithe, seiceálaithe gramadaí, seiceálaithe stíle agus, faoi láthair, is iad na huirlisí próiseála teanga is fairsinge atá in úsáid iad. Bíonn siad réamhshuite sna cláir phróiseála focal do mhórtheangacha an domhain, m.sh., MS Word, OpenOffice, s.rl., chomh maith le feidhmchláir ríomhphoist mar MS Outlook.

Tá an Seiceálaí Litrithe Gaelspell ar fáil go saoráideach agus é saor in aisce, é sin forbartha ag Ollscoil St. Louis, Missouri . Is féidir é a úsáid ar ardán ar bith (Linux, Mac, nó Windows) agus tá pacáistí ar féidir iad a shuiteáil le hardáin foinse oscailte mar Mozilla Firefox, Mozilla Thunderbird, OpenOffice nó LibreOffice. Tá sé curtha in oiriúnt do bhogearraí eile chomh maith, mar shampla, Microsoft Word agus Mac OS X . Is sampla maith é seo den tslí ar féidir pacáiste oscailte a úsáid ar shlite nach raibh i gceist nuair a bhíothas á bhforbairt.

Forbraíodh Litreoir Gaeilge Microsoft Word i gColáiste na Tríonóide, Baile Átha Cliath agus tá fáil air i gcomhair Microsoft Office le húsáid in Microsoft Word. Is féidir é a íoslódáil ó Microsoft Irish Proofing Tools . Níor tugadh é seo suas chun dáta, áfach, ó 2002 agus is gá é seo a dhéanamh anois .

An Gramadóir : seiceálaí gramadaí foinse oscailte riailbhunaithe is ea é atá ar fáil saor in aisce. Forbraíodh é seo in Ollscoil St. Louis, Missouri chomh maith. Úsáideann sé clibeáil ranna cainte agus 2,800 riail a leanann patrún chun na hearráidí is coitianta sa Ghaeilge a aithint. Tá sé seo comhcheangailte le bogearraí cosúil le Microsoft Word (via na pacáistí 'Ceart' agus 'Anois' atá ar díol ag Cruinneog) agus LibreOffice (comhfhiontar idir Dúrud agus COGG) .

Teorainn amháin atá leis ná nach ndéanann sé anailís dhomhain teangeolaíochta ar an gcomhréir agus mar thoradh air sin ní thagann sé ar roinnt de na hearráidí agus uaireanta bíonn toradh bréagdheimhneach air. D'fhéadfaí é a fheabhsú trí uirlisí próiseála nádúrtha teanga níos sofaisticiúla a chur i bhfeidhm, m.sh., smutáin riailbhunaithe nó staitisticiúla, nó parsálaí spleáchais (féach Caibidil 3).

Tarraingíonn Uirlisí Meastacháin Inléiteachta aird an scríbhneora ar cé chomh maith is a luífeadh an sliocht leis an léitheoir ó thaobh tuisceana de. Cuireann sé leis na seiceálaithe gramadaí agus litrithe agus tugann sé moltaí faoi conas an téacs a dhéanamh níos soiléire agus níos fusa le léamh. Tá a leithéid ar fáil le fada an lá don Bhéarla [1], m.sh., Grammarly [2], ach níl sé ar fáil don Ghaeilge fós. Tá buntaighde tosaithe chomh maith a rachaidh i dtreo uirlis meastacháin inléiteachta a fhorbairt [3]. Faraor, faoi mar atá d'fhormhór na dteicneolaíochtaí urlabhra agus teanga, is ar an mBéarla den chuid is mó atá na fothacaíochtaí teangeolaíochta bunaithe. Tá na háiseanna a luadh i gCuid 1 riachtanach chun freastal ar chás na Gaeilge.

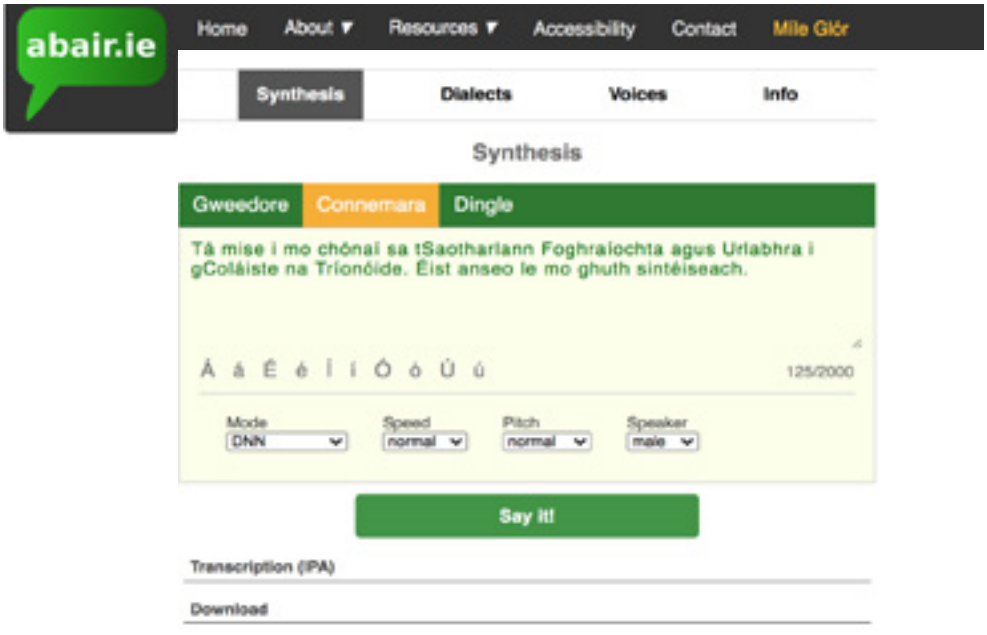
Tuarthéacs do Ghléasanna Soghluaiste. Úsáidtear Adaptxt ar fhóin phóca chun cur le hionchur téacs agus é a dhéanamh níos tapúla agus níos cruinne. Déanann sé tuar ar an bhfocal atá á scríobh chomh maith leis an bhfocal a leanfaidh é agus bíonn sé de shíor ag dul i ngleic le stór focal agus le stíl an scríbhneora. Le nascanna.com is féidir Adapttxt le foclóir Gaeilge a shocrú mar mhéarchlár príomha chun teacht i dtír ar thuarthéacs agus seiceálaí litrithe don Ghaeilge.

Breiseán Foclóra. Tugann breiseán foclóra an deis don té a bhfuil leathanach gréasáin á léamh aige dul chuig an bhfoclóir gan an leathanach ar a bhfuil sé a fhágáil. Tá suíomhanna gréasáin fhoclóirí na Gaeilge ag cur na mbreiseán seo ar fáil de réir a chéile. Léiríteari bhFíor 13.1 an breiseán a ghabhann le teanglann.ie agus é in úsáid leis an nuachtán ar líne tuairisc.ie. Is féidir an uirlis luachmhar seo a neartú trí fheidhmeanna breise a chur léi, mar shampla, an cumas foirmeacha infhillte a láimhseáil (féach Caibidil 2.5.1).

Figure 13.1: breiseán teanglann.ie ar shuíomh tuairisc.ie

13.4.2 Áiseanna Caintbhunaithe

Tá bealaí éagsúla anois ann chun téacs scríofa ar líne a léamh amach. Ar leathanach gréasáin ABAIR, www.abair.ie, tá fáil ar réimse guthanna sintéiseacha do na trí mhórchanúint agus úsáidtear an leathanach seo freisin mar thairseach faisnéise do thaighde ar theicneolaíocht urlabhra na Gaeilge. Úsáidtear an suíomh gréasáin seo go forleathan in Éirinn agus thar lear. Tá tuilleadh feidhmeanna agus bogearraí ar fáil ar an tairseach seo atá dírithe ar rochtain do dhaoine atá faoi mhíchumas agus ar chúrsaí oideachais.



Fíor 13.2: suíomh gréasáin www.abair.ie

Tá aip Android ag ABAIR freisin agus leis seo is féidir téacsanna a léamh amach ar fhóin chliste agus ar tháibléid . Canúint Chonamara amháin atá ar fáil ar an aip faoi láthair ach beifear ag cur leis sin go luath.

Tá breiseán ag ABAIR freisin a forbraíodh chun é a dhéanamh níos fusa teacht ar fhoirm labhartha de théacs ar aon saghas gléis ríomhaireachta. Is féidir leis an úsáideoir é a íoslódáil ar a bhrabhsálaí agus é a chur ag gníomhú nuair a bhíonn téacs Gaeilge á léamh aige. Áis eile is ea an léitheoir gréasáin agus is féidir leis na forbróirí í seo a chur le haon leathanach gréasáin . Sainléitheoir ar leith atá ann lena mbíonn an rogha chanúna agus an luas léitheoireachta faoi smacht an úsáideora agus is féidir focail aonaracha a aibhsiú de réir mar atá siad á léamh amach (féach Caibidil 8). I gcás iontrálacha ón bhfoclóir nuair a scríobhtar ~ (tilde) in áit an cheannfhocail líontar isteach an fhréamh atá ar lár nuair a léitear an focal amach. Mar shampla, dá mbeadh '~faidh' le feiceáil do 'tóg' san aimsir fháistineach, chuirfeadh an breiseán an ceannfhocal leis agus chloisfeadh an t-úsáideoir 'tógfaidh'.

Tá tábhacht ar leith ag baint leis an léitheoir gréasáin do chomhlachtaí poiblí a bhfuil sé de dhualgas orthu faoin dlí eolas a chur ar fáil trí Ghaeilge agus chun rochtain a chinntiú don uile dhuine, daoine faoi mhíchumas san áireamh (féach an tAcht Míchumais, 2005 §26(1) (a)). Tá cur síos ar roinnt feidhmchlár atá bunaithe ar chaint shintéiseach le fáil i gCaibidlí 8, 14 agus 15. Tabhair faoi deara, áfach, go bhfuil cuid mhaith acu seo úsáideach do phobal ginearálta na Gaeilge. Mar shampla, beidh fáilteoirí i bhfoirm róbait, bunaithe ar chórais chainte (Caibidil 10) in úsáid níos minice in áiteanna poiblí mar ghailearaithe eolaíochta, iarsmalanna agus a leithéid eile.

13.4.3 Logánú na bhFeidhmchlár is Coitianta

Is éard atá i gceist le logánú ná earraí nó ábhair a chur in oiriúint do cheantair nó do mhargaí ar leith. Tá dhá ghné ag baint leis: ar dtús é a aistriú isteach sa teanga logánta agus ansin an t-ábhar a chur in oiriúint do chultúr agus nósanna an cheantair – cúrsaí polaitíochta, dáta, am, córas airgeadais, dathanna, fuaimeanna agus riachtanais dlí san áireamh.

Ceann de na haidhmeanna a théann leis an bPlean Digiteach ná an Ghaeilge a chur in úsáid i ngach gné den saol. Forbraíodh cuid mhór den saol digiteach le go mbeadh sé oiriúnach don Bhéarla, rud atá difriúil le comhthéacs na Gaeilge. Má bhíonn leagan Gaeilge de na hardáin dhigiteacha atá in úsáid againn go rialta ar fáil cuirfidh sé sin le líon na réimsí saoil inar féidir an Ghaeilge a úsáid go nádúrtha agus le gradam na Gaeilge mar theanga.
An Ríomhacadamh . Eagraíocht dheonach í seo ina bhfuil forbróirí, aistritheoirí agus teicneolaithe a bhfuil sé d'aidhm acu bogearraí a chur ar fáil do phobal na Gaeilge ionas gur féidir leo feidhmiú ar líne go hiomlán trí mheán na Gaeilge. Go dtí seo tá logánú déanta acu ar Gmail, Twitter, Mozilla Firefox, OpenOffice, LibreOffice, Code.org, Scratch, WhatsApp, Skype agus go leor eile feidhmchlár, suíomhanna greasáin agus cluichí. Tá treoir stíle curtha ar fáil acu do logánú bogearraí Gaeilge agus iarrtar ar bhaill é seo a leanúint ionas go mbeidh comhsheasmhacht acu nuair a thagann sé go téarmaíocht, gramadach, réim theanga, srl. do na hearraí go léir atá ar fáil trí Ghaeilge .

Tá aistritheoirí proifisiúnta faighte ag gnólachtaí eile idirnáisiúnta cosúil le Microsoft agus Airbnb chun a gcuid seirbhísí a chur in oiriúint don Ghaeilge. Tabhair faoi deara, áfach, go bhfuil deireadh curtha leis an tacaíocht don leagan logánta de MS Office ó MS Office 2010 i leith. Maidir le bogearraí foinse oscailte, tá tacaíocht leanúnach d'earraí Gaeilge curtha ar fáil ag OpenOffice.

Sampla maith de logánú ina bhfuil an t-ábhar curtha in oiriúint don chomhthéacs áitiúil is ea an físchluiche Minecraft. Rinneadh an t-aistriúchán trí shluafhoinsiú a chur ar bun agus cuireadh é seo le hábhar tíreolaíochta (féach Caibidil 4) a tógadh ó Shuirbhéireacht Ordanáis Éireann le go bhféadfadh na himreoirí Éire fhíorúil a shiúl . Tá gnólachtaí eile mór le rá a tháirgíonn cluichí ríomhaireachta, ar nós Romero Games, ag tosú ar logánú a dhéanamh ar a gcluichí do phobal na Gaeilge .

Mar a dúradh i gCaibidil 1, ní féidir teicneolaíocht don Ghaeilge a chruthú díreach trí aistriúchán agus mionathruithe a dhéanamh ar na cinn Bhéarla. Faoi mar a tharlaíonn má bhímid ag forbairt teicneolaíochta don Ghaeilge (féach Cuid I agus Cuid II) ní mór dúinn struchtúr cuí teangeolaíochta a chur san áireamh, chomh maith leis an gcomhthéacs ina bhfuil lucht a húsáidte, nuair a bhíonn logánú á dhéanamh ar earraí agus ar áiseanna atá cruthaithe do mhargaí idirnáisiúnta. Mar sin, má bhímid chun logánú a dhéanamh ar bhogearraí le haghaidh fóin phóca, mar shampla, níor mhór uirlis tuarthéacs a d'aithneodh focail na Gaeilge, agus a leanfadh struchtúir teangeolaíochta na Gaeilge, a chruthú.
Cuirfidh na háiseanna, na huirlisí agus na teicneolaíochtaí ar fad atá molta sa Phlean Digiteach seo le gnéithe éagsúla den logánú. Tá an

meaisínaistriúchán thar a bheith tábhachtach agus cuirfidh sé go mór leis an logánú amach anseo.

13.5 Moltaí do na Blianta atá Romhainn

Tiocfaidh rudaí nach bhfeicimid faoi láthair chun cinn le himeacht aimsire agus tiocfaidh daoine cruthaitheacha i dtír orthu. Spreagfaidh gach céim bhreise den Phlean Digiteach seo nuálaíochtaí nach bhfuil ar ár súile faoi láthair. Agus ardáin agus feidhmchláir á dtógáil de réir mar atá an bhunteicneolaíocht á forbairt beifear níos cinnte gur féidir leis an dá thaobh suí le chéile le go mbeidh siad oiriúnach don chomhthéacs a bhfuil siad ceaptha dó.

Tá an méid seo thíos á mholadh ionas go mbeidh na torthaí ar an bPlean Digiteach ag freagairt don phobal chun go mbainfidh siad tairbhe astu:

i. Ba cheart go dtabharfaí tús áite do thógáil na bhfeidhmchlár de réir riachtanais an phobail.
ii. Is páirtnéir lárnach é pobal na Gaeilge agus ba cheart go mbeadh áit lárnach acu i bhforbairt na bhfeidhmchlár ón gcéad lá. Is féidir ionchur a bheith acu ag leibhéil éagsúla, chun cabhrú le tosaíochtaí a shocrú agus tástáil a dhéanamh ar earraí de réir mar atá siad ag teacht chun cinn. (Tá tuilleadh faoi seo sa Chonclúid).
iii. Beidh gá ag na grúpaí forbartha le tacaíocht ionas go mbeidh na hearraí tarraingteach le féachaint orthu, éasca le suiteáil, coinnithe i gceart agus uasdátaithe le go mbeidh siad comhoiriúnaithe d'athruithe sna córais oibriúcháin (m.sh. uasghrádú ar Windows), sna teangacha ríomhchlárúcháin (m.sh. uasghrádú ar Python) agus a leithéid eile. Moltar saoráid leanúnach tacaíochta d'úsáideoirí freisin.
iv. Beidh gá le huirlisí comhéadanacha agus feidhmchláir a bheith comhoiriúnach mar uirlisí tríú páirtí d'ardáin choitianta cosúil le Microsoft 365. Beidh gá freisin le tacaíocht chothabhála agus teicneolaíochta. Ba dheis mhór é seo do ghnólachtaí chun a dtacaíocht do phobail áitiúla a léiriú mar chuid dá bhFreagracht Shóisialta Chorparáideach.
v. Caithfidh bunuirlisí teanga agus feidhmeanna don phobal a bheith leabaithe mar dhlúthchuid de phacáistí digiteacha a chuireann earnáil tionsclaíochta ríomhaireacht na hÉireann ar fáil.
vi. Tacaíocht ón Rialtas lena chinntiú go mbeidh na huirlisí agus na feidhmeanna ina ndlúthchuid den bhunfhearas a chuirtear ar fáil don earnáil phoiblí – státseirbhísigh, scoileanna agus institiúidí tríú leibhéal san áireamh.
vii. Tacaíocht ón Rialtas chun an teicneolaíocht nua a chur os comhair an phobail agus daoine a spreagadh chun í a úsáid. Chomh maith leis sin d'fhéadfadh an Rialtas spreagadh a thabhairt do ghnólachtaí móra teicneolaíochta ról na Gaeilge a aithint agus tacaíocht a thabhairt di mar chéadteanga náisiúnta na tíre.

Tagairtí

[1] Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. 1975.

[2] Grammarly Inc. Grammarly [Internet]. 2014. Available from: https://github.com/grammarly

[3] Ó Meachair, M.J. The Creation and Complexity Analysis of a Corpus of Educational Materials in Irish (EdaGA). Trinity College, Dublin; 2019.

Caibidil 14
Cúrsaí Oideachais agus Foghlaim Ríomhchuidithe Teanga (FRT)

14.1 Réamhrá

Ceann de na réimsí is tairbhí agus is nithiúla maidir le maoiniú a chur ar fáil do Phlean Digiteach don Ghaeilge ná múineadh na teanga. Is féidir an Ghaeilge a chur ina dlúthchuid de shíorfhás theicneolaíocht na linne seo agus an fhís a léirigh an Roinn Oideachais agus Scileanna a thabhairt chun blátha – is é sin "cumas na dteicneolaíochtaí digiteacha a bhaint amach le feabhas a chur ar theagasc, foghlaim agus measúnú" [1]. Cuirfidh na moltaí ón bPlean Digiteach leis an bhfís seo ar go leor bealaí éagsúla. Úsáidfear na huirlisí is nua-aimseartha chun anailís chainníochtúil a dhéanamh ar shonraí eimpíreacha d'fhonn athspléachadh a dhéanamh ar struchtúir na Gaeilge agus ar an modh a sealbhíonn daoine an Ghaeilge mar chéad nó mar dhara teanga. Cuirfidh sé seo bunús eolaíochta le dearadh curaclaim chomh maith le hábhair mhúinteoireachta agus ábhair mheastóireachta. Ina theannta sin cuirfidh na teicneolaíochtaí lárnacha atá sa phlean leis an úsáid spreagúil is féidir a bhaint as cláir éagsúla don fhoghlaim ríomhchuidithe teanga (FRT) chun forbairt réabhlóideach a dhéanamh ar fhoghlaim agus ar mhúineadh na Gaeilge. Cuideoidh siad freisin le húsáid na Gaeilge i réimsí oideachasúla eile.

Leathnaíonn FRT ó chláir shimplí go cinn atá an-chasta. Ar láimh amháin is féidir cláir a chruthú atá réasunta simplí a chabhróidh le sainscileanna ar leith a fhorbairt, m.sh. mionchluichí le hamharcíomhánna tarraingteacha éifeachtacha. Ar an taobh eile den scála tá feidhmeanna ann atá bunaithe ar intleacht shaorga chun ardáin choimpléascacha a chruthú le húsáid na teicneolaíochta is déanaí. Ina measc siúd tá ardáin ilmhódacha, chaintbhunaithe, idirghníomhacha ina n-úsáidtear eilimintí de phróiseáil teanga nádúrtha – áit a mbíonn ról gníomhach ag an bhfoghlaimeoir i gcluiche atá saindeartha agus struchtúrtha nó ardán a thógann an foghlaimeoir isteach i réaltacht fhíorúil thumach ina mbíonn an t-idirphlé ar fad trí Ghaeilge. Ina leithéid sin d'ardáin baintear úsáid as teicneolaíocht shofaisticiúil a forbraíodh do chluichí ríomhaireachta agus nasctar é seo le heolas ar an teanga ar mhaithe le cuspóirí foghlama teanga. Aithnítear go bhféadfadh tionchar ollmhór a bheith acu seo ar chomhthéacs fhoghlaim na Gaeilge agus tá béim ar leith orthu sa chaibidil seo.

Chun ardáin chasta mar atá luaite thuas a fhorbairt beidh gá le foirne ildisciplíneacha – dearthóirí comhéadan úsáideoirí/ríomhchluichí agus forbróirí bogearraí ina measc (féach 14.5 thíos). Thar aon ní eile, áfach, is é an t-ionchur ó theagascóirí agus teangeolaithe a bhfuil saineolas acu ar FRT an ceann is tábhachtaí. Tá gá freisin le hionchur ó mhúinteoirí agus ó oideachasóirí eile lena chinntiú go bhfuil na hardáin ag teacht le (i) prionsabail mhaithe mhúinteoireachta, (ii) tuiscint ar struchtúr agus ar shealbhú na teanga agus (iii) soiléireacht faoi na húsáideoirí agus a gcuid sainriachtanas siúd. Sa dianiarracht seo beifear ag brath go mór ar an teangeolaíocht agus ar eolas eile atá luaite i gCuid I, anailís ar chorpais foghlaimeoirí san áireamh (féach Caibidil 2).

14.2 Cén fáth a bhfuil FRT tábhachtach i gcás na Gaeilge?

Ó bunaíodh an Stát aithníodh go raibh oideachas lárnach i gcaomhnú na Gaeilge mar theanga labhartha in Éirinn. Cé go bhfuil an Ghaeilge éigeantach mar ábhar scoile suas go dtí an Ardteist (rud a fhágann go bhfuil pobal mór foghlaimeoirí ann) agus cé go bhfuil muintir na hÉireann fabhrach don Ghaeilge [2] ní chiallaíonn sé sin go mbíonn siad cíocrach chun foghlama ná go mbíonn torthaí fónta ar a gcuid iarrachtaí [3]. Tá an caighdeán ag titim le blianta beaga anuas [4] agus is é an trua é go mbíonn cuid mhaith daltaí ann nach mbíonn ábalta comhrá bunúsach a dhéanamh trí Ghaeilge ar fhágáil na scoile dóibh tar éis trí bliana déag a chaitheamh á foghlaim.

Is é an fhadhb is mó atá ann ná nach mbíonn go leor teagmhála ag formhór na bhfoghlaimeoirí leis an teanga ná le daoine a bhfuil Gaeilge ó dhúchas acu,

toisc nach teanga phobail í an Ghaeilge ach amháin i gceantair Ghaeltachta. Tá a laghad sin daoine ann a bhfuil Gaeilge ó dhúchas acu a bheadh ina n-eiseamláirí ag foghlaimeoirí agus a laghad sin comhthéacsanna ann ina dtarlaíonn cumarsáid nádúrtha trí Ghaeilge (an comhthéacs is fearr i gcomhair foghlama) go mbíonn bac ar fhoghlaim éifeachtach. Is minic a bhíonn múinteoirí, ar foghlaimeoirí iad féin, míchompordach faoina gcumas féin agus is minic a bhraitheann siad leo féin gan tacaíocht agus slánú na Gaeilge ag luí go trom orthu [5]. Braitheann tuismitheoirí faoi mhíbhuntáiste freisin agus iad ag iarraidh cabhrú le hobair bhaile, inimircigh go háirithe.

Aontaíonn go leor léirmheastóirí gurb é fuaimniú nádúrtha na Gaeilge an ghné is lú a bhfuil rath air i sealbhú na teanga. Tá córas casta fuaimnithe ag an nGaeilge atá an-difriúil le córas fuaimnithe an Bhéarla (Caibidil 2.4), tá an tras-scríbhinn ón bhfuaim go dtí an focal scríofa casta (Caibidil 2.6.3) agus cuireann an dá rud sin le deacracht san fhuaimniú agus sa litearthacht. Go hidéalach, tá gá le cur chuige sa rang a chuirfeadh ar chumas na bhfoghlaimeoirí greim a fháil go nádúrtha agus i ngan fhios dóibh féin ar na fuaimeanna agus ar an gcomhfhreagairt atá eatarthu seo agus an focal scríofa (tuilleadh eolais le fáil i gCaibidil 7). Tá gnéithe eile de struchtúr na teanga ann freisin, moirfeolaíocht agus comhréir mar shampla, atá difriúil go maith leis na gnéithe sin sa Bhéarla agus cruthaíonn siad deacrachtaí ar leith (Caibidil 2.5, 2.6).

Ar an iomlán tá easpa áiseanna ar fáil don Ghaeilge i gcomparáid le teangacha eile. Tá go leor de na háiseanna sin as dáta agus is beag a dhéanann siad chun dul i bhfeidhm ar na foghlaimeoirí óga, a bhfuil taithí acu ar an teicneolaíocht is nua-aimseartha sna mórtheangacha, agus fágann sé sin díspreagtha iad le drochbhlas ar an nGaeilge. Ó thaobh na nGaelscoileanna de tá easpa áiseanna iontu do mhúineadh ábhar eile seachas an Ghaeilge. Sa Straitéis 20 Bliain don Ghaeilge, 2010-2030 a d'fhoilsigh Rialtas na hÉireann [6], tá sé mar aidhm ann roinnt ábhar seachas an Ghaeilge a mhúineadh trí Ghaeilge sna gnáth-bhunscoileanna (ábhar agus teanga a mhúineadh i dteannta a chéile nó Foghlaim Chomhtháite Ábhar agus Teangacha (CLIL)). Tá sé feicthe go bhfuil an cur chuige sin an-éifeachtach i dtíortha eile [7] ach beidh sé deacair é a chur i bhfeidhm anseo de cheal ábhar múinteoireachta. Seo arís ceann de na réimsí ina bhféadfadh forbairt áiseanna FRT cur go mór leis an gcás.

Ar ndóigh, ní neamh ar talamh é FRT, ach nuair a chuirtear é i dteannta réimsí eile den Phlean Digiteach seo agus na hacmhainní cuí curtha leis tugann sé deis iontach cuid de na fadhbanna atá luaite thuas a shárú. Is mór an chabhair é don fhoghlaimeoir áis a bheith aige chun teacht ar an bhfuaimniú atá ag téacs láithreach. Má thugtar é seo ar aghaidh go dtí an chéad leibhéal eile, le hardáin chaintchumasaithe atá idirghníomhach, ilmhódach agus bunaithe ar chluichí, chuirfí ar chumas an fhoghlaimeora dul isteach sa saol fíorúil agus feidhmiú le guthanna atá ar aon dul leis an nGaeilge nádúrtha. Ina theannta sin, ba cheart go spreagfadh an spraoi a ghabhfadh leis an gcluiche na foghlaimeoirí dul i ngleic leis na ceachtanna teanga agus chun coinneáil leo. De réir taighde a rinneadh le déanaí spreagann cluichí oideachasúla ilmhódacha le saol fíorúil (a bhfuil cur síos gairid déanta orthu thíos) foghlaimeoirí chun rannpháirtíocht ghníomhach a ghlacadh san fhiontar [8], [9]. An rud is tarraingtí faoi na hardáin seo ná go gcuirtear béim iontu ar an bhfocal labhartha agus ar scileannna cumarsáide, ach gan an ghramadach ná an cruinneas scríbhneoireachta a bheith fágtha in áit na leathphingne.

Tá sé tábhachtach go dtugann na hardáin FRT seo taithí thumthach iomlánaíoch don fhoghlaimeoir a bheadh mar chrann taca don oiliúint uilíoch teanga. Cé go bhféadfadh cluichí áirithe a bheith dírithe ar phointí áirithe gramadaí agus feidhm a bhaint as áiseanna próiseála teanga nádúrtha, san am céanna bheadh na foghlaimeoirí ag éisteacht agus ag labhairt mar chuid den chluiche agus chabhródh sé seo leo le fuaimniú agus le tuiscint na teanga chomh maith leis na scileanna gramadaí. Tá sé seo ag teacht le tuairimíocht an lae inniu faoi shealbhú teanga: moltar cur chuige uilíoch cothromaithe ina múintear gnéithe éagsúla den teanga i dteannta a chéile, seachas léamh, scríobh, tuiscint agus labhairt a mhúineadh ceann ar cheann [9, 10] [13].

Tá an teicneolaíocht dhigiteach ar tí a bheith i lár an aonaigh ó thaobh na múinteoireachta de agus caithfidh an Ghaeilge a bheith mar chuid den réabhlóid seo nó fágfar ar an trá fholamh í. Tá sé ríthábhachtach go ndéanfaí teicneolaíochtaí na hurlabhra agus na teanga a chónascadh le cluichí oideachasúla ionas go gcuirfear le héifeacht na foghlama/na múinteoireachta. I ndeireadh na dála is í an aidhm atá ann ná foghlaimeoirí a spreagadh agus a chur ar a súile dóibh go bhfuil an Ghaeilge beo beathach agus in ann a bheith ina dlúthchuid den saol digiteach.

14.3 Cé dó FRT ?

Daltaí Scoile: D'fhéadfadh FRT teagasc/foghlaim na Gaeilge a athrú ó bhonn. Tréimhse spreagúil atá anseo don Ghaeilge: tá teicneolaíochtaí na hurlabhra agus na teanga atá luaite sa Phlean Digiteach (Cuid II) ag teacht chun cinn agus nuair a chuirtear iad seo i dteannta na

teangeolaíochta agus na n-áiseanna atá luaite i gCuid I, tá an fhéidearthacht ann ardáin chliste chomhtháite a fhorbairt sa ghearrthréimhse atá romhainn amach. Nuair a bheidh ardáin ann a bheidh idirghníomhach, tumthach agus cluiche-bhunaithe, athróidh siad an taithí a bheidh ag daltaí scoile ó bhonn. Ní féidir beag is fiú a dhéanamh den tionchar a bheidh aige seo ar na tuairim is 700,000 dalta atá ag foghlaim na Gaeilge sa chóras scolaíochta. Beidh gá le hardáin a fhorbairt a bheidh oiriúnach do gach leibhéal, réamhscoileanna san áireamh.

Tuismitheoirí: Chabhródh ardáin mar seo le tuismitheoirí agus iad ag cabhrú lena bpáistí le hobair bhaile, mar shampla le fuaimniú na bhfocal sa luathléitheoireacht.

Foghlaimeoirí Fásta: Bíonn díograiseoirí fásta ag iarraidh cur lena gcumas Gaeilge. Tá suim mhór sa Ghaeilge i dtíortha thar lear mar is léir ón tuairim is 60,000 foghlaimeoir i 138 tír éagsúil atá cláraithe don chúrsa Fáilte ar Líne - Fáilte Online [11]. Tá tarraingt dhochreidte ar an nGaeilge ar an ardán Duolingo agus tuairiscítear le déanaí go bhfuil suas le 1 mhilliún duine ag gabháil do cheachtanna Gaeilge gach seachtain. Tá ardú tagtha ar líon na gcúrsaí atá á dtairiscint ag ollscoileanna thar lear agus tá foghlaimeoirí le fáil ar fud na cruinne, m.sh., sna Stáit Aontaithe, i gCeanada, sa Bhreatain, sa Rúis agus ar mhór-roinn na hEorpa.

Foghlaimeoirí Féinrialaithe: Tá FRT thar a bheith oiriúnach d'fhoghlaimeoirí féinrialaithe [12] agus tá sé seo thar a bheith tábhachtach do dhaoine fásta agus d'fhoghlaimeoirí thar lear ach, ar ndóigh, chabhródh sé le gach saghas foghlaimeora.

Daoine le Gaeilge ó Dhúchas: Cé go bhfuil an plé sa chaibidil seo dírithe ar fhoghlaimeoirí go príomha, d'fhéadfadh teicneolaíochtaí na teanga agus na hurlabhra cabhrú le cainteoirí dúchasacha go díreach nó go hindíreach, mar is léir ón taithí sna mórtheangacha. Bheadh aidhmeanna agus cuspóirí difriúla acu seo, áfach.

Múinteoirí: Níl sé i gceist go dtiocfadh FRT in áit na múinteoirí ach go mbeadh sé ina chabhair dóibh trí theicneolaíocht agus áiseanna a chur ina lámha. Tugann FRT an deis do mhúinteoirí ábhar struchtúrtha ar ardchaighdeán a chur ar fáil ionas go mbeidh siad ábalta freastal ar leibhéil éagsúla chumais sna ranganna agus ábhair atá sainoiriúnaithe do riachtanais ar leith a chur os comhair a ndaltaí i dteanga nádúrtha labhartha ar deacair teacht uirthi go minic i gcomhthéacs an ranga. Mar atá pléite thíos (§14.5), tá ról lárnach ag múinteoirí i bhforbairt agus i seachadta FRT. Anuas air sin, tá ról ag lucht forbartha curaclaim chun FRT a shníomh isteach mar dhlúthchuid de shiollabais. Beidh sé riachtanach sa todhchaí go mbeidh an dá ghrúpa sin lárnach i ndearadh agus i bhforbairt ábhar in FRT. Beidh gá freisin le seachadadh éifeachtach FRT sa rang a thabhairt faoi scáth na gcúrsaí oiliúna do mhúinteoirí nua agus beidh siad seo ábalta ionchur a bheith acu san fhorbairt.

An Roinn Oideachais: Is gá don Roinn ról gníomhach a ghlacadh i gcur chun cinn FRT.

Oideolaíocht mar Chroí: Cé gur mó páirtnéir atá san fhiontar, braithfidh fiúntas FRT ar chaighdeán na teangeolaíochta agus na hoideolaíochta a bheidh ag dul leis. Caithfear súil an-ghéar a choimeád ar an gcaighdeán seo mar is cuma cé chomh tarraingteach is atá an pacáiste, mura mbíonn an t-ábhar ar chaighdeán ard ann tarraingeoidh sé droch-cháil ar an gcur chuige atá á mholadh anseo.

14.4 A bhfuil déanta i leith na Gaeilge go dtí seo

Áiseanna digiteacha atá ar fáil faoi láthair
Is beag sainardán FRT atá ar fáil don Ghaeilge faoi láthair, ardáin atá ábalta teacht i dtír ar theicneolaíocht na cainte agus na teanga. Ta roinnt mhaith ábhair eile ar fáil ar líne, áfach, chomh maith le háiseanna próiseála teanga nádúrtha agus corpais a thabharfaidh treoir d'ábhair theagaisc agus do dhearadh siollabais amach anseo.

Ábhair Foghlama ar Líne: Is léir ón éileamh atá ar ábhair tharraingteacha le haghaidh foghlaim teangacha an gá atá le creataí ilmheánacha do dhaltaí scoile. I measc na n-ábhar atá anois ann tá Bua na Cainte [13] agus Abair Liom [14] do scoileanna Béarla; Séideán Sí [15] do Ghaelscoileanna; Cód na Gaeilge [X], Féasta Focal [16] agus Foghlaí Focal [17] d'fhoghlaimeoirí i dTuaisceart Éireann chomh maith le hábhair eile ar líne ([18], [19]) agus roinnt feidhmeanna oideachasúla ([20] – [22]). Tá treoracha luachmhara do na háiseanna ar fad atá ar fáil ar na suíomhanna gréasáin atá ag an gComhairle um Oideachas Gaeltachta agus Gaelscolaíochta (COGG) agus ag Gaeloideachas ([23], [24]).

Áiseanna Próiseála Teanga Nádúrtha: Tá áiseanna luachmhara PTN (féach Caibidlí 4 agus 5) in úsáid go forleathan ag foghlaimeoirí cheana féin. Ina measc seo tá: seiceálaithe litrithe [X], seiceálaithe gramadaí [X,Y], foclóirí agus áiseanna téarmaíochta [X,Y,Z]. Taobh thiar díobh seo tá áiseanna corpais [X], anailíseoirí/gineadóirí moirfeolaíochta [X], clibeálaithe ranna cainte [25], crannbhainc agus parsálaithe [XNasc ag teastáil anseo don chuid cuí den bPlean Digiteach].

Corpais Foghlaimeoirí: Tá corpais foghlaimeoirí atá dírithe ar chúrsaí oideachais ag teacht chun cinn (Caibidil 3) agus beidh siad in úsáid do dhearadh curaclaim, do mheasúnú agus d'ábhar múinteoireachta. Is é an aidhm atá ag cuid acu seo ná measúnú a dhéanamh ar ghramadach agus ar stór focal an fhoghlaimeora chun iad a rangú de réir na leibhéal atá ag dul le Comhchreat Tagartha na hEorpa um Theangacha (CEFR) [26]. Tá anailís leanúnach á déanamh ar chorpais na dtéacsleabhar [X] atá in úsáid agus beidh grádú á dhéanamh orthu seo de réir choimpléascacht a stóir focal agus a gcomhréire.

Ardáin FRT: Fréamhshamhlacha Reatha
D'fhéadfadh an Plean Digiteach athrú ó bhonn a chur ar an tslí a dtéimid i mbun an Ghaeilge a fhoghlaim. Táimid ag dréim le hardáin chasta chliste a thiocfaidh i dtír ar theicneolaíochtaí na teanga agus na hurlabhra agus a chuirfidh an fhoghraíocht agus an teangeolaíocht i gcroílár na múinteoireachta ar shlí a bheidh tarraingteach d'aos óg na ré digití seo.

Athchló ar Ardáin Reatha: má tá ardáin ar fáil a bhfuil cuid de na gnéithe inmhianaithe ag roinnt leo, mar shampla, scéalta fíorúla tarraingteacha nó tréithe eile ar féidir iad a leasú, d'fhéadfaí iad seo a athchóiriú agus a chur in oiriúint d'aidhm nua. Is é atá i gceist anseo ná creatlach de scéal nó de chluiche a thógail agus é seo a leasú chun ábhar teagaisc a chruthú a bheadh oiriúnach d'fhoghlaimeoirí ar leith. Sampla de seo is ea Digichaint , atá ina chluiche réamhshamhaltach, idirghníomhach, ilmhodúlach atá curtha in oiriúint d'fhoghlaimeoirí. Sa chluiche seo bíonn an foghlaimeoir ag measc adh le scata carachtar in óstán fíorúil d'fhonn fadhb a réiteach [27]. Forbraíodh ábhar an phlota ó bhonn, agus coimeádtar ann na buntréithe oiriúnaithe ionas go mbíonn éifeacht ag na roghanna a dhéanann an foghlaimeoir ar leibhéal na teanga a chuirtear os a chomhair. Dá bharr seo d'fhéadfaí an cluiche seo a úsáid i ranganna ina bhfuil leibhéil éagsúla chumais. Réamhshamhail dhifriúil is ea Fáilte go TCD , áis ina dtugtar an deis d'fhoghlaimeoirí cúléisteacht le grúpa ata ag labhairt as Gaeilge sa Chearnóg Thosaigh i gColáiste na Tríonóide [8], [28]. Tá an comhrá leagtha amach d'fhoghlaimeoirí a bhfuil caighdeán maith Gaeilge acu agus é bunaithe ar shiollabas na hArdteiste. Úsáidtear guthanna sintéiseacha de chuid ABAIR sa dá ardán seo.

Ardáin Nua do Shaincheisteanna a bhaineann leis an nGaeilge: is trí ardáin nua a fhorbairt, ina ndíreofar ar shaincheisteanna don Ghaeilge, a thiocfar ar fhíoréifeacht FRT. Ní mór an fhoghraíocht agus an teangeolaíocht chuí a bheith leabaithe iontu seo (Cuid I) agus iad curtha in oiriúint d'ábhar an tseomra ranga. Tabharfaidh sé seo deis ardáin chliste idirghníomhacha a fhorbairt ar bhealach nach féidir a dhéanamh trí athchóiriú a dhéanamh ar chinn atá réamhthógtha.

Sampla de chluiche nuathógtha is ea Lón don Leon , cluiche ina ndírítear ar thuiscint a fhorbairt ar chodarsnacht chasta chonsain na Gaeilge (Caibidil 2.4) chomh maith le hiontuiscintí ar an tslí a n-aistrítear é seo isteach i gcóras scríbhneoireachta, eadhon forbairt na foghraíochta (féach Caibidil 2.6.3) don luathléitheoireacht. Tá gá le hábhar nuachruthaithe, graificí, scéalta agus ceol san áireamh, bunaithe ar na haidhmeanna teangeolaíochta a bhfuiltear ag díriú orthu. I gcóras mar seo nasctar caint nádúrtha le hamhráin agus guthanna sintéiseacha agus le háiseanna teangeolaíochta atá sa chóras téacs-go-hurlabhra, m.sh. sainrialacha canúintbhunaithe litreacha-go-fuaimeanna (féach Caibidil 8).

Sampla eile de seo is ea An Scéalaí , ardán FRT cliste atá dírithe ar shainghnéithe de struchtúr na Gaeilge. Ina lár seo tá guthanna sintéiseacha agus áiseanna próiseála teanga nádúrtha. Ceann de na buntréithe a théann leis seo ná go ndírítear ann ar an teanga a fhorbairt ina hiomláine ionas go bhforbrófar scileanna scríbhneoireachta agus scileanna léitheoireachta i dteannta scileanna cluastuisceana agus scileanna labhartha. Sa leagan reatha den ardán is é an buntasc atá ann ná scéal a scríobh, ach ón tús éisteann an scríbhneoir le guth sintéiseach á léamh. Tugann an phroféisteacht seo seans don fhoghlaimeoir earráidí coitianta a thabhairt faoi deara agus daingníonn sé an comhcheangal idir an leagan labhartha agus an leagan scríofa. Úsáidtear téacsanna scríofa, téacsanna labhartha agus léaráidí chun an foghlaimeoir a spreagadh chun na hearráidí a cheartú mura n-éiríonn leis an gcluas na hearráidí a aimsiú. Faoi láthair, tá an bhéim ar dheilbhíocht agus ar fhóineolaíocht na Gaeilge ach táthar ag cur leis seo de réir mar a thagann na háiseanna atá luaite i gCuid I chun cinn. Tá an t-ardán modúlach agus tiocfaidh sé i dtír ar na teicneolaíochtaí atá luaite i gCuid II de réir mar a thagann siad chun blátha, an córas comhrá ríomhbhunaithe (Caibidil 10), go háirithe.

Is áis taighde do FRT cliste é an t-ardán freisin. Coimeádtar ionchur na bhfoghlaimeoirí lena n-iarrachtaí scríofa agus a n-iarrachtaí labhartha chomh maith le taifead ar a gcuid iarrachtaí ar earráidí a cheartú. Leis seo tá corpas foghlaimeoirí á thógáil agus tá anailís á déanamh air chun tosaíochtaí teangeolaíochta a chruthú agus chun aidhmeanna múinteoireachta a chéimniú d'ardáin nua amach anseo. Tugann sé léargas an-mhaith freisin ar phróiseas na foghlama teanga.

Idirghníomhaíocht Chumarsáideach an Eochair: tá na hardáin seo ar fad idirghníomhach, ach bíonn buntáistí móra ann d'fhoghlaimeoirí teanga nuair a bhraitheann siad go bhfuil siad ag déanamh fíorchumarsáide. Léiríonn an t-ardán píolótach Taidhgín (féach Caibidil 10) gur féidir leis na teicneolaíochtaí atá ann cheana féin, fiú, an chuma a chruthú gur fíorpháirtnéirí comhrá atá

iontu, fiú más moncaí atá ag caint! Léiríonn an taighde a rinneadh i scoileanna go raibh na daltaí an-ghafa leis, fiú amháin nuair a bhí an réimse topaicí a raibh sé de chumas ag an moncaí labhairt fúthu an-teoranta. Tá an chaint ríthábhachtach ina leithéid seo d'ardán: ní féidir comhrá neamhiata a chruthú gan caint shintéiseach nuair nach féidir ionchur an fhoghlaimeora a thomhas roimh ré agus, mar sin, ní féidir réamhthaifeadadh a dhéanamh ar thaobh amháin den chomhrá. Cuireann próiseáil teanga nádúrtha le 'clisteacht' an ardáin seo (féach Caibidil 10). Is comhtháthú é de chaint chliste agus teicneolaíochtaí teanga agus tógann sé seo aird na ndaltaí ar shlí a chuireann spraoi leis an bhfoghlaim.

14.5 An Ród atá Romhainn: ag Tógáil ar FRT

Is disciplín feidhmiúil é FRT agus baineann sé saineolas agus uirlisí saothair ó réimsí éagsúla léinn. Léirítear i bhFíor 14.1 thíos an tslí a dtógann an lárfhoireann taighdeoirí na teicneolaíochtaí, idir shean agus nua, atá leagtha amach i gCuid II agus na háiseanna atá leagtha amach i gCuid IIe chéile. Téann an lárfhoireann seo i gcomhairle le múinteoirí, le daltaí agus le foghlaimeoirí eile chun comhpháirtíocht a chruthú idir na daoine leasmhara go léir. Tá léiriú san Fhíor thíos ar an tslí a dtagann na daoine agus na hacmhainní eile le chéile chun ardáin Ghaeilge don todhchaí a chruthú.

Fíor 14.1: Comhpháirteanna d'ardáin FRT don Ghaeilge

(1). An lárfhoireann taighde/forbartha FRT a dhéanann an dearadh, an fhorbairt agus an tástáil ar na hardáin. Nascann siad an t-ionchur ó na réimsí éagsúla agus oibríonn siad lámh ar lámh le múinteoirí agus le foghlaimeoirí. I gcroílár an fhiontair cruthaíonn an/na comhordaitheoir(í) ábhar teagaisc/foghlama do na hardáin agus nascann siad é seo leis an teicneolaíocht agus an modh teagaisc. Bíonn gá freisin le ríomhchláraitheoirí agus le dearthóirí chun comhéadan cuí a chur ar an ardán. Ós é go mbíonn forluí idir na scileanna teicniúla, ba cheart go mbeadh an fhoireann forbartha ag obair lámh ar lámh leo siúd atá ag obair ar an mbunteicneolaíocht a luíonn taobh thiar de.

(2). Teicneolaíochtaí Urlabhra agus Teanga (Cuid II). De réir mar a thagann na bunteicneolaíochtaí ata luaite i gCuid II chun cinn, cuirfidh siad ar ár gcumas réimse leathan ardán a chruthú. Ós rud é go bhfuil tábhacht ar leith ag baint le feidhmeanna oideachasúla i gcomhthéacs na Gaeilge, tá sé tábhachtach go dtabharfaí tús áite don fhorbairt teicneolaíochta sin a bheadh ag teastáil don oideachas. Mar atá léirithe thuas tá tábhacht ar leith le córais chomhrá ríomhbhunaithe. Tá gá acu seo le haithint cainte, rud nach bhfuil ar fáil go forleathan fós, ach a bhfuiltear ag obair air (Caibidil 9).

Cé go bhfuil an obair ar chaint shintéiseach tagtha píosa fada, chun riar ceart a dhéanamh ar na feidhmchláir FRT atá beartaithe, beifear ag súil le breis solúbthachta a chur leis na guthanna chun iad a chur in oiriúint do charachtair éagsúla a bheadh sna cluichí agus chun go mbeadh mothucháin sa phrosóid chun iad a dhéanamh níos inchreidte agus iad ag labhairt (Caibidil 8). Ní gá d'fhorbróirí FRT fanacht, áfach, go dtí go mbeidh na bunteicneolaíochtaí forbartha go hiomlán. Mar a léirigh an taithí le Taidhgín, fiú mura bhfuil ann ach córas comhrá ríomhbhunaithe an-bhunúsach, gan aithint cainte uaithoibríoch, fós féin is féidir leis a bheith an-tarraingteach mar áis foghlama. Caithfear tús áite a thabhairt anois do ghuthanna leanaí a fhorbairt don tsintéis agus don aithint cainte.

Chabhródh córas meaisínaistriúcháin, a bheadh ar ardchaighdeán, le cúrsaí oideachais freisin, leis na Gaelscoileanna go háirithe. Faoi láthair, mar shampla, dá mbeadh dalta ag déanamh taighde ar líne i gcomhair tionscadail, is trí Bhéarla is mó a chaithfeadh sé an obair a dhéanamh. D'fhéadfaí é sin a athrú leis an teicneolaíocht. Beidh áiseanna próiseála teanga nádúrtha agus foclóirí leabaithe i gceartlár na n-ardán foghlama sa todhchaí.

(3). Saineolas ar Theanga agus Urlabhra agus Acmhainní (Cuid I). Tá an saineolas agus na hacmhainní atá á bhforbairt ón taighde atá leagtha amach i gCuid I i gcroílár na bhfeidhmchlár FRT cliste eolasbhunaithe. Siod é mar a thiontaítear taighde ar theangeolaíocht agus ar fhoghraíocht isteach i bhfoirm atá cabhrach do mhúineadh agus d'fhoghlaim teangacha. Mar atá ráite thuas, is féidir le gréasáin shéimeantacha

focal (Caibidil 4.5) a bheith an-úsáideach chun pointí gramadaí a fhoghlaim go neamh-chomhfhiosach nuair atá na gréasáin sin leabaithe in ardán idirghníomhach [9]. Is féidir áiseanna guthbhunaithe a shamhlú chun codarsnachtaí fuaime na Gaeilge a mhúineadh, chomh maith le múnlaí prosóide le haiseolas físeach (Caibidil 7). Is minic a dhéantar dearmad ar shainfhuaimniú na Gaeilge ach d'fhéadfaí cluichí a shamhlú ina ndéanfaí na scileanna sin níos inghlactha don fhoghlaimeoir. Is gá, ar ndóigh, go mbeadh lucht taighde sa teangeolaíocht agus san fhoghraíocht ag obair i gcomhar leis an lárfhoireann FRT agus le saineolaithe teagaisc. Tabharfaidh an obair seo gnéithe de struchtúr na Gaeilge chun solais, gnéithe nár cuireadh ar a súile do mhúinteoirí ná d'fhoghlaimeoirí na Gaeilge cheana (m.sh. múnlaí prosóide, comhréire agus séimeantaice). Ba cheart na hacmhainní seo ar fad a fhorbairt ar shlí a chuirfeadh lena gcumas freastal ar riachtanais oideachasúla agus freastal ar dhaoine faoi mhíchumas ó thaobh na rochtana de.

(4). Oideolaíocht roimh Theicneolaíocht: Múinteoirí, Dearthóirí Curaclaim, Saineolaithe i Sealbhú an Dara Teanga & Oideolaithe. Tá sé tábhachtach gan an botún a dhéanamh in FRT agus an teicneolaíocht a chur chun tosaigh ar an oideolaíocht. Agus é ag cur síos ar fhorbairt FRT tráth, shamhlaigh Underwood é mar 'letting computer whizzes first explore how the computer can do something particularly well and then design the program to take advantage of that' [29, lch. 83]. Moltar a mhalairt sa Phlean seo, is é sin go bhfaigheadh an oideolaíocht tosaíocht ar an teicneolaíocht. Is í an fhoghlaim teanga an sprioc agus úsáidtear an teicneolaíocht chun an sprioc sin a bhaint amach. Caithfidh na haidhmeanna oideachasúla a bheith bunaithe ar chomhthéacs agus ar riachtanais na bhfoghlaimeoirí agus ní mór an teicneolaíocht a chur in oiriúint chun freastal orthu sin.

Is í an oideolaíocht bunaidhm FRT agus is ar chaighdeán an ábhair múinteoireachta a dhéanfar breithmheas ar FRT i ndeireadh na dála. Dá bharr seo is den riachtanas é go mbeadh na forbróirí ábhar ar fhoireann FRT ag obair lámh ar lámh leis na múinteoirí agus leis na dearthóirí curaclaim. Tá gá freisin le saineolas ar shealbhú an dara teanga, sealbhú na Gaeilge go háirithe. Tá sé tábhachtach nach leagtar cuid de na tuiscintí atá ann faoi shealbhú an Bhéarla anuas ar an nGaeilge gan iniúchadh a dhéanamh ar an gceist.

(5). Na húsáideoirí: Is don fhoghlaimeoir é FRT agus mar sin caithfidh sé/sí a bheith ina d(h)lúthpháirtí i ngach céim den fhorbairt. Is é an teist is mó a bheidh ar an ardán ná an meas atá ag na foghlaimeoirí air agus mar a thógann siad chucu féin é. Is den tábhacht é go mbunófaí gréasáin de mhúinteoirí agus d'fhoghlaimeoirí a d'oibreodh i gcomhar leis an lárfhoireann chun moltaí a thabhairt agus chun measúnú a dhéanamh ar fhorbairtí de réir mar a thagann siad chun tosaigh.

Tagairtí

[1] Department of Education and Skills, "Digital Strategy for Schools 2015-2020: Enhancing Teaching, Learning and Assessment," 2015.

[2] M. Mac Gréil and F. Rhatigan, "The Irish language and the Irish people: Report on the attitudes towards, competence in and use of the Irish language in the Republic of Ireland in 2007-08," National University of Ireland Maynooth, 2009.

[3] L. Murtagh, "Retention and attrition of Irish as a second language: A longitudinal study of general and communicative proficiency in Irish among second level school leavers and the influence of instructional background, language use and attitude/motivation variable," University of Groningen, Amsterdam, 2003.

[4] J. Harris, P. Forde, P. Archer, S. Nic Fhearaile, and M. O'Gorman, "Irish in primary schools: Long-term national trends in achievement," Dublin: Stationery Office, 2006.

[5] C. Dunne, "Becoming a Teacher of Irish: The Evolution of Beliefs, Attitudes and Role Perceptions," Trinity College Dublin, 2015.

[6] _ Rialtas na hÉireann, "Straitéis 20 Bliain don Ghaeilge 2010-2030: 20-year strategy for the Irish language 2010-2030," Baile Átha Cliath: Oifig an tSoláthair, 2010.

[7] D. Coyle, P. Hood, and D. Marsh, CLIL: content and language integrated learning. Cambridge University Press, 2010.

[8] N. Ní Chiaráin, "Text-to-Speech Synthesis in Computer-Assisted Language Learning for Irish: Development and Evaluation," (Doctoral thesis, CLCS, Trinity College, Dublin), 2014.

[9] N. Ní Chiaráin and A. Ní Chasaide, "The Potential of Text-to-Speech Synthesis in Computer-Assisted Language Learning: A Minority Language Perspective," in Recent Tools for Computer- and Mobile-Assisted Foreign Language Learning, A. Andujar, Ed. Hershey, PA: IGI Global, 2020, pp. 149–169.

[10] M. Pressley and R. L. Allington, Reading instruction that works: the case for balanced teaching, Fourth edi. The Guilford Press, 2015.

[11] O. C. B. Á. Cliath, "Fáilte Ar Líne - Fáilte Online," 2020. [Online]. Available: www.failteonline.ie/ga/.

[12] J. Egbert, E. Hanson-Smith, and C. Chao, "Foundations for Teaching and Learning (Caibidil 1, revised)," in CALL environments: Research, practice, and critical issues (2nd edition), 2nd editio., J. Egbert and E. Hanson-Smith, Eds. Alexandria, VA: TESOL, 2007, pp. 19–28.

[13] S. de Brún and M. Ní Fhatharta, "Bua na Cainte," EDCO, 2014. .

[14] Folens, "Abair Liom," 2015. [Online]. Available: www.folens.ie/programmes/abair-liom/products.

[15] A. Gúm, T. D. E. and Skills, E. and A. Council for the Curriculum, and F. na Gaeilge, "Séideán Sí," Online material available for Junior and Senior Infants, 2015. [Online]. Available: www.seideansi.ie.

[16] N. I. CCEA, Council for the Curriculum, Examinations and Assesment, "Féasta Focal." [Online]. Available: http://ccea.org.uk/curriculum/gaeloideachas/eochairchéimeanna_1_2/réimsí_foghlama/teanga_agus_litearthacht/féasta_focal.

[17] N. I. Council for the Curriculum, Examinations and Assessment, "Foghlaí Focal." [Online]. Available: http://ccea.org.uk/curriculum/gaeloideachas/foghlaí_focal.

[18] N. Tarrant, "Léigh Leat," 2014. [Online]. Available: leighleat.com.

[19] "subh.ie," funding from COGG, 2013. [Online]. Available: www.subh.ie.

[20] E. Ní Arrachtáin, "Lingua App," Oideas Gael, 2015. [Online]. Available: www.lingua-app.ie.

[21] "Abair Leat Oide," Coláiste Lurgan, 2015. [Online]. Available: https://play.google.com/store/apps/details?id=air.AbairleatOideMobile&hl=en.

[22] E. Ní Ghallachobhair, "Enjoy Irish," funding from COGG, 2011. [Online]. Available: www.enjoyirish.ie/#anchor.

[23] COGG, "Resource Database," An Chomhairle um Oideachas Gaeltachta agus Gaelscolaíochta, 2020. [Online]. Available: https://www.cogg.ie/bunchar-aiseanna/#?keywords=&subject=&types=&levels=&page=1&language=ga.

[24] Gaeloideachas, "Gaeloideachas: guth don Oideachas lán-Ghaeilge agus Gaeltachta," 2020. [Online]. Available: https://gaeloideachas.ie/.

[25] E. Uí Dhonnchadha, Natural language processing tools: Developing a part-of-speech tagger and partial dependency parser for Irish. Saarbrücken: LAP Lambert Academic Publishing, 2010.

[26] Council of Europe, "Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)," 2020. [Online]. Available: https://www.coe.int/en/web/common-european-framework-reference-languages/home.

[27] N. Ní Chiaráin and A. Ní Chasaide, "The Digichaint interactive game as a virtual learning environment for Irish," in CALL communities and culture – short papers from EUROCALL 2016, S. Papadima-Sophocleous, L. Bradley, and S. Thouësny, Eds. Research-publishing.net, 2016, pp. 330–336.

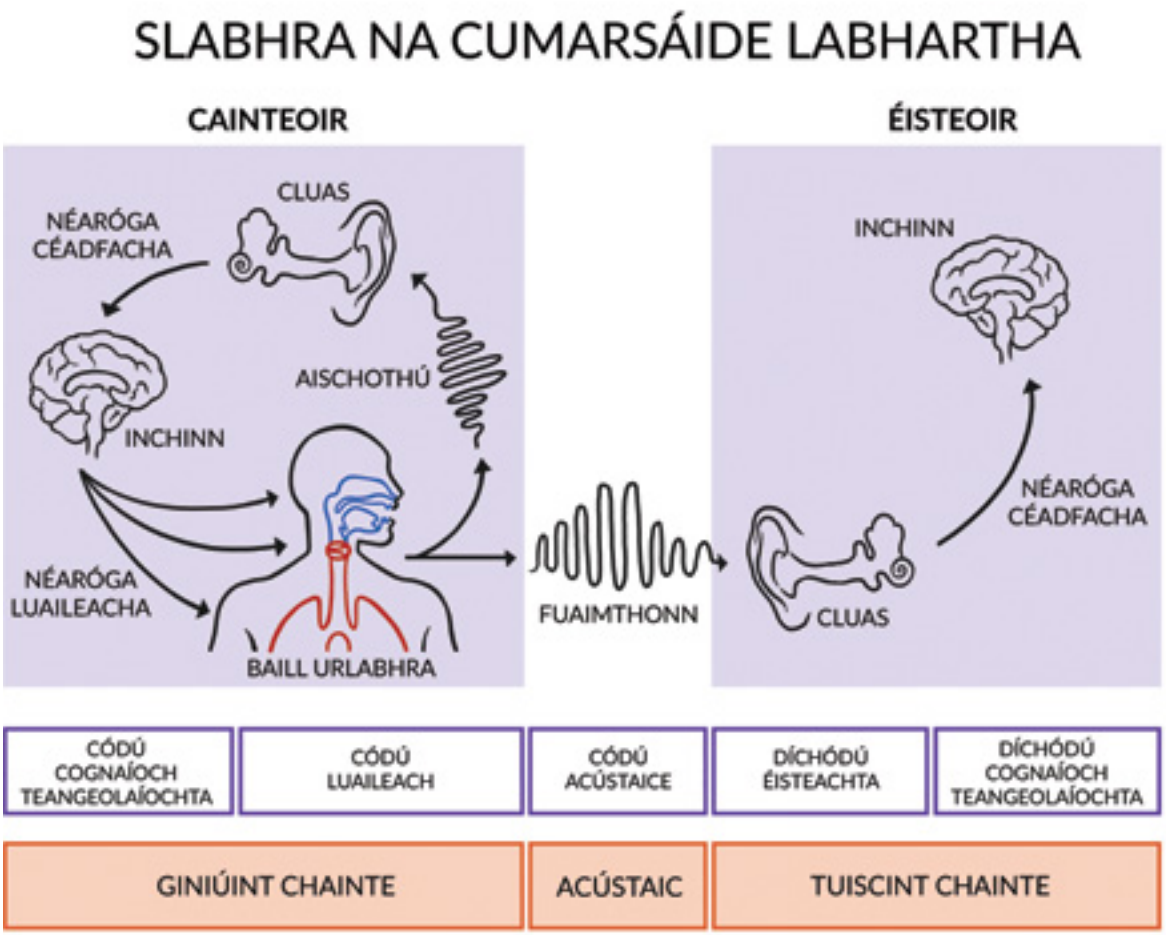
[28] N. Ní Chiaráin and A. Ní Chasaide, "Evaluating Synthetic Speech in an Irish CALL Application: Influences of predisposition and of the holistic environment," in SLaTE 2015: 6th Workshop on Speech and Language Technologies in Education, 2015, pp. 149–154.

[29] J. Underwood, Linguistics, computers and the language teacher: a communicative approach. Rowley, MA: Newbury House, 1984.

Caibidil 15
Áiseanna: Míchumas agus Rochtain

15.1 Cén rud é?

Tá teicneolaíochtaí urlabhra agus teanga fíorthábhachtach dóibh siúd faoi mhíchumas. Sa chumarsáid labhartha tá slabhra de chéimeanna [1], léirithe i bhFíor 15.1, a théann ó ghiniúint na cainte (an cainteoir) > an tonnchruth acústaice > aithint na cainte (an t-éisteoir).



Figiúr 15.1: Slabhra na Cumarsáide Labhartha (bunaithe ar [1]).

Giniúint na Cainte

- Códú cognaíoch/teangeolaíochta: aistrítear smaointe an chainteora i bhfoirm aonad teanga agus seoltar teachtaireachtaí chuig na baill urlabhra chun na haonaid seo a chur i crích.
- Códú luaileach: le teachtaireachtaí ó na néaróga luaileacha, bogann na baill urlabhra: aistríonn téada an ghutha (an glór-fhoinse) an anáil ina fhuaim, a scagtar le hurlabhraíocht an bhéil (Caibidil 7 & §2.4).

Códú Acústaice

- Sa tonnchruth acústaice tá 'códú na teanga' le patrúin na foinse agus scagadh na fuaime.

Tuiscint na Cainte

- Díchódú éisteachta, airíonn an chluas an fhuaim agus iompraíonn na néaróga céadfacha an cód fuaime chuig an inchinn.
- Díchódú cognaíoch/teangeolaíochta: aithníonn an t-éisteoir na mórthréithe acústaice a léiríonn 'cód na teanga' as a dtuigtear ciall na cainte.

Corruair, briseann ar an slabhra ag pointe ar leith, mar gheall ar fhadhbanna cognaíocha, luaile nó éisteachta. Is féidir go minic le teicneolaíocht na hurlabhra an bhris a shárú: mar shampla, má tá tuiscint na teanga ag duine ach nach bhfuil sí/sé ábalta labhairt, is féidir le gléas sintéiseach an bhearna a líonadh agus bealach a chur ar fáil don chumarsáid. Baintear úsáid fhorleathan as an teanga scríofa fosta, ach is mó duine le fadhbanna radharc, luaile nó disléicse nach bhfuil teacht acu ar an fhocal scríofa. Is féidir leis an teicneolaíocht na fadhbanna rochtana seo a shárú fosta.

Meastar go hidirnáisiúnta go bhfuil míchumas ar líon ard daoine: duine as ceathrar sna Stáit Aontaithe [2], agus i suirbhé Eorpach [3] meastar 17.5% dóibh siúd thar aois 15. Nuair a bhaineann an míchumas le briseadh i slabhra na cumarsáide nó le deacrachtaí téacs a léamh, a scríobh nó a fheiceáil, cuireann sé isteach ar gach gné saoil. Is mó gairm bheatha nach féidir tabhairt faoi, fiú i gcás daoine atá sárcháilithe. Meastar go bhfuil 75% de dhaoine faoi mhíchumas radhairc dífhostaithe san Eoraip [4], agus is mar an gcéanna an scéal in Éirinn [5]. Tá rochtain ar an oideachas agus ar eolas i bhfad níos deacra, agus is minic srian ar an leibhéal oideachais a bhaintear amach. Tá eisiamh sóisialta go minic mar thoradh ar aon bhris i slabhra na cumarsáide.

Tugann áiseanna urlabhra agus teanga bealach le bris na cumarsáide a thrasnú, le teacht ar oideachas, agus le páirt níos iomláine a ghlacadh sa saol gairmiúil agus sóisialta. Ar an drochuair, is beag áis chumarsáide atá ann don Ghaeilge. Go rómhinic molann lucht leighis agus múinteoirí do thuismitheoirí a bhfuil páiste faoi mhíchumas acu gan an Ghaeilge a fhoghlaim ar scoil, gan freastal ar Ghaelscoil, agus fiú in amanna, gan an Ghaeilge a labhairt sa bhaile. Go rómhinic is míthuiscint atá taobh thiar de chomhairle mar seo, agus easpa eolais ag lucht leighis/teagaisc ar na buntáistí a bhaineann leis an dátheangachas do pháistí óga, rud atá léirithe sa taighde arís is arís eile. Tá sé fíorthábhachtach go mbeidh na teicneolaíochtaí cumarsáide Gaeilge ann dóibh siúd faoi mhíchumas, a cheadóidh rochtain ar oideachas trí Ghaeilge agus ar shaol gairmiúil agus sóisialta mar bhaill de phobal na Ghaeilge. Mar sin, i bhforbairt áiseanna sa Phlean Digiteach, moltar go mbeidh béim faoi leith ar na riachtanais acu siúd faoi mhíchumas, agus go ndeartar gach áis ó thús chun gur féidir le gach duine iad a úsáid – rud atá mar bhun choincheap sa ghluaiseacht "Design for All" [6]. Tá na hacmhainní a bhaineann le Cuid I den Phlean tábhachtach fosta chun freastal mar is cóir orthu siúd faoi mhíchumas agus tá géarghá leo do mhórán áiseanna a bheidh dírithe orthu.

Tá mórán réimsí lena dtacaíonn teicneolaíochtaí agus acmhainní digiteacha. Sa chaibidil seo, ní féidir clúdach a dhéanamh orthu, ach pléitear go gairid na himpleachtaí dóibh siúd faoi mhíchumas labhartha (§15.2), faoi mhíchumas radhairc (§15.3), le fadhbanna litearthachta, mar an disléicse (§15.4), agus impleachtaí don teiripe teanga agus urlabhra (§15.5). Tugtar moltaí ginearálta in §15.6.

15.2 Áiseanna teicneolaíochta cúnta dóibh siúd faoi mhíchumas urlabhra

15.2.1 Cé dó iad agus cén tábhacht a bhaineann leo?

I gcás míchumais mar an phairilis cheirbreach nó an t-uathachas, is féidir nach n-éiríonn le duine caint intuigthe a bhaint amach. I gcásanna eile, cailltear cumas na cainte mar gheall ar ghalar mar stróc, galar Parkinson, scléaróis iolrach, galar néarón luadrach nó ailse na dtéad gutha.

Mar a léirigh cás Stephen Hawkins, fiú le galar néarón luadrach, le háis sintéise cainte is féidir páirt iomlán a ghlacadh sa saol, ainneoin an mhíchumais labhartha. Dóibh siúd atá in ann an méarchlár a láimhseáil, is cuidiú mór an tuarthéacs agus seiceálaí litrithe agus gramadaí. Ach do mhórán daoine faoi mhíchumas luaile, ní féidir an méarchlár a smachtú agus tá gléasanna speisialta cúnta de dhíth (córais AAC) a cheadaíonn dóibh an teachtaireacht don chóras sintéise a chumadh. Is aip ar ríomhaire táibléid atá i gceist go minic: roghnaíonn an t-úsáideoir sraith pictiúr, siombal nó litreacha, a

aistrítear ina n-abairtí le labhairt amach. Nuair nach féidir na lámha a bhogadh ar chor ar bith, is féidir téacs a chumadh ar bhealach eile, m.sh. trí do shúile a leagadh ar na pictiúir, srl.

Tá taighde agus dul chun cinn tapa ag tarlú sa réimse seo. Mar shampla, dóibh siúd ar féidir leo labhairt ach nach bhfuil a gcuid cainte intuigthe, tá taighde ar fhéidearthachtaí na haitheanta cainte á dhéanamh, le féachaint an bhféadfaí córas a thraenáil a d'aithneodh an chaint, ionas go bhféadfaí an teachtaireacht a labhairt amach le glór sintéiseach [7]. Bheadh a leithéid d'áis ag tógáil ar na modhanna a úsáidtear chun barr fheabhais a chur ar chórais aitheanta cainte do chainteoir ar leith (Caibidil 9.5). Dóibh siúd a labhraíonn trí ghlór sintéiseach, tá sé ríthábhachtach go bhfeileann an glór don úsáideoir (féach thíos agus Caibidil 8). Tá forbairt déanta sa CSTR i nDún Éideann ar mhodhanna nuálacha le hoiriúnú pearsanta a dhéanamh ar an ghlór sintéiseach, ag baint úsáid as corpais ollmhóra a bailíodh don aithint cainte. Rinneadh forbairt fosta ansin ar bhealach le córais phearsanta shintéiseacha a thógáil do dhaoine ar a bhfuil an cumas labhartha ag teip, ag baint úsáid as taifeadadh dá nglór féin, sula gcaillfear é.

Tá féidearthachtaí as cuimse ann don todhchaí, agus bainfear úsáid níos mó agus níos mó as abhatáranna/botaí le cumas cainte. Tá forbairt á déanamh a cheadódh do dhuine le siondróm iniata (atá in ann teachtaireacht a chumadh ina cheann, ach nach bhfuil bealach ar bith aige chun é a chur in iúl) gléas sintéiseach a smachtú trí theachtaireachtaí (códú luaile) ó choirtéis luadrach na hintinne.

15.2.2 Cad é atá ar fáil go dtí seo?

An tsintéis chainte agus áiseanna cúnta (AAC): tá an tsintéis chainte ar fáil in aisce do na trí mhórchanúint . Faoi láthair níl áiseanna cúnta (AAC) ann dóibh siúd nach féidir leo méarchlár a láimhseáil, ach tá tús curtha ina threo ag grúpa ABAIR i gColáiste na Tríonóide . Tá cur síos ar chruthú coinceapa d'áis AAC i [31].

Áiseanna le haithint cainte: níl a leithéid ann go fóill, ach tá féidearthachtaí anois ag teacht chun cinn, faoi mar a dhéantar dul chun cinn ar na córais aitheanta cainte (Caibidil 9).

15.2.3 Riachtanais don todhchaí

Áiseanna cúnta AAC le sintéis chainte: dóibh siúd atá in ann an méarchlár a úsáid, tá gá le tuarthéacs chun cumadh na dtéacsanna don chóras sintéise a bhrostú. Dóibh siúd nach féidir leo méarchlár a úsáid, tá uirlisí cúnta AAC de dhíth go práinneach – uirlísí a bheidh seasmhach agus iomlán oirúnach don lucht úsáide. Ní forbairt shimplí ar chomhéadain don chórais sintéiseacha atá i gceist anseo, ná aistriú go Gaeilge ar na focail do na siombail i gcóras AAC don Bhéarla. Beidh buntaighde de dhíth ar ghiniúint na cainte ó choincheap (ó phictiúir/ó shiombail), ar bhealach a luíonn le struchtúr na teanga (atá an-difriúil ón Bhéarla). Agus toisc go mbeidh na gléasanna seo in úsáid i gcomhthéacs dátheangach, caithfear an córas AAC a dhearadh chun gur féidir abairtí sa dá theanga a ghiniúint ar bhealach atá éasca don úsáideoir.

An glór sintéiseach agus féiniúlacht an úsáideora: is gné thábhachtach dár bhféiniúlacht ár nglór. Nuair is do ghlór féin an glór sintéiseach, tá sé tábhachtach go luíonn sé le d'fhéiniúlacht. Sa chéad áit caithfidh sé a bheith in oiriúint do chanúint, inscne agus aois an úsáideora – agus tá tábhacht faoi leith le glórtha páistí. Sa bhreis air sin, tá smachtú ar cháilíochtaí an ghlóir ríthábhachtach ar dhá bhealach (féach Caibidlí 7 agus 8):

- Do mhionathruithe do cháilíochtaí bunúsacha an ghlóir shintéisigh ionas go n-oireann sé do phearsantacht agus d'fhéiniúlacht an té atá á úsáid (glór bog, glór láidir, srl);
- Chun meon agus mothúcháin a chur in iúl nuair atá muid ag comhrá, athraíonn muid cáilíochtaí an ghlóir de shíor (prosóid an ghlóir) chun mothúcháin a nochtadh (glór feargach, cairdiúil, brónach, srl). Is cúis mhór díomá é don té faoi mhíchumas 'labhairt' gan mothúcháin sa ghlór.

Áiseanna le haithint cainte: faoi mar a thagann an aithint cainte chun cinn, beidh go leor bealaí inar féidir é a chur in úsáid. Beidh sé tábhachtach díriú ar bhealaí inar féidir an córas a chur in oiriúint do dhuine faoi leith – fiú le cainteoirí nach bhfuil iomlán intuigthe - chun a gcuid cainte a aistriú i bhfoirm téacs, nó labhartha le glór sintéiseach.

Do gach gné taighde agus forbartha, beidh ionchur de dhíth ó lucht úsáide na n-áiseanna, óna gcúramóirí, ó theiripeoirí agus ó na heagrais a bhaineann le míchumas.

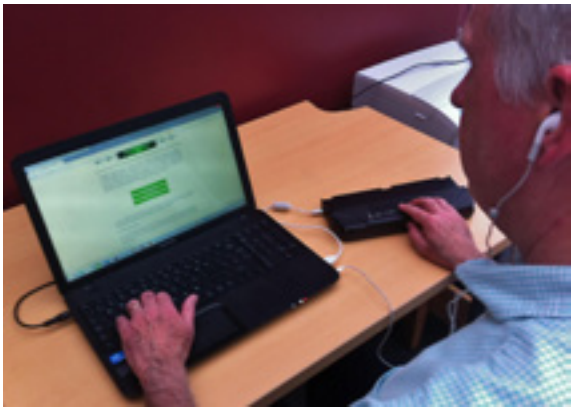
15.3 Áiseanna dóibh siúd faoi mhíchumas radhairc

15.3.1 Cé dó iad agus cén tábhacht a bhaineann leo?

Nuair nach bhfuil teacht ar an fhocal scríofa, tá impleachtaí móra don oideachas, don ghairm agus don tsaol shóisialta. De réir dhaonáireamh na bliana 2016, tá 54,810 duine dall nó faoi mhíchumas radhairc in Éirinn, agus is páistí scoile 4,701 acu (5-18 bliain)

[5]. Tá deacrachtaí ar leith ag páistí Gaeltachta agus Gaelscolaíochta agus is minic an-bhrú orthu a gcuid scolaíochta a dhéanamh trí mheán an Bhéarla.

Is riachtanas bunúsach an léitheoir scáileáin le haschur cainte sintéiseach le ceadú dóibh siúd gan radharc nó ar bheagán radhairc éisteacht le téacs ar ríomhaire agus bogearraí an ríomhaire a láimhseáil go hiomlán ó chluais. Tá closleabhair úsáideach fosta, go háirithe i gcás téacsleabhair do pháistí scoile. Tá áiseanna don deachtú bunaithe ar an aithint cainte tábhachtach fosta, chun gur féidir téacs a chumadh ó bhéal.



Figiúr 15.2: An Léitheoir scáileáin NVDA le breiseán ABAIR in úsáid le córas Braille Liblouis

15.3.2 Cad é atá ar fáil go dtí seo?

Léitheoir scáileáin: tá breiseán le glórtha sintéiseacha Gaeilge ar fáil don léitheoir scáileáin ar fhoinse oscailte NVDA [10]. Is mar fhreagra d'achainí práinneacha ó thuismitheoirí le páistí faoi mhíchumas radhairc a rinneadh an fhorbairt seo sa tionscadal ABAIR, le tacaíocht ón NCBI , ó COGG agus ó speisialtóirí múinteoireachta a rinne tástáil air le daltaí scoile. Is féidir leis an úsáideoir an chanúint agus luas na cainte a roghnú agus a aibhsiú agus méadú a dhéanamh ar an téacs de réir mar a léitear amach é. Is féidir an córas Liblouis Braille a chur ag obair leis, agus aschur labhartha agus Braille a fháil ag an am céanna, faoi mar a léirítear i bhFíor 15.2.

Leabhair Ilmheán Daisy [11]: tá sraith leabhar ilmheán le haschur cainte do pháistí scoile faoi mhíchumas radhairc ullmhaithe le glórtha sintéiseacha ABAIR agus ar fáil ó Childvision Ireland.

Áiseanna don deachtú le haithint cainte: níl a leithéid ar fáil go fóill.

15.3.3 Riachtanais don Todhchaí

Áiseanna don deachtú: beidh gá le córais a bheidh in ann réimse leathan cainteoirí a aithint, le canúintí difriúla, agus tá tábhacht ar leith ag baint le glórtha páistí.

Léitheoirí Scáileáin: cé go bhfuil aschur cainte agus Braille Gaeilge le fail leis an bhreiseán don léitheoir scáileáin NVDA, is gá comhéadain do na léitheoirí scáileáin eile atá in úsáid go forleathan.

Dearadh na n-áiseanna do chách: is iomaí áis ghinearálta a bheadh úsáideach do dhaoine le breacamharc – m.sh. an léitheoir ar líne lenar féidir aibhsiú agus méadú a dhéanamh ar an téacs, cluichí oideachasúla agus áiseanna do theagasc na litearthachta (Caibidil 8, 13, 14). Mar atá luaite thuas, nuair a dheartar cibé áis, is gá a chinntiú ó thús go mbeidh sí inúsáidte acu sin ar bheagán radhairc – agus acu siúd faoi mhíchumais éagsúla.

Míchumas coimpléascach: sa chás go bhfuil níos mó ná míchumas radhairc i gceist, beidh gá fosta leis na comhéadain agus leis na háiseanna cúnta a phléitear in §15.2.

I ngach gné forbartha beidh comhpháirtíocht de dhíth le húsáideoirí, le speisialtóirí múinteoireachta, le lucht teiripe agus leis na heagrais a bhaineann le míchumas radhairc.

15.4 Áiseanna agus acmhainní dóibh siúd le fadhbanna litearthachta mar an disléicse

15.4.1 Cé dó iad agus cén tábhacht a bhaineann leo?

Meastar go bhfuil an disléicse ar thart ar 10% de dhaonra na hÉireann [12]. Ós rud é go bhfuil thart ar milliún mac léinn lánaimseartha sa tír [13], fágann sé sin go bhfuil fadhbanna disléicse ag thart ar 100,000 dalta. Cé go bhfaigheann go leor acu díolúine ó scrúduithe, tá go leor eile ar mian leo a dteanga náisiúnta a fhoghlaim. Tá oideachas trí mheán na Gaeilge á chur ar pháistí na Gaeltachta agus ar líon maith páistí lasmuigh den Ghaeltacht. Is léir ón taighde idirnáisiúnta gur mór an buntáiste i bhforbairt an linbh a bheith dátheangach go hóg. Ar mhaithe le cothromaíocht san oideachas agus cothromaíocht teanga tá sé ríthábhachtach go mbeadh sé de cheart ag gach uile dalta an teanga a fhoghlaim, freastal ar Ghaelscoil, agus go mbeadh an tacaíocht chuí ar fáil.

Baintear úsáid as áiseanna teicneolaíochta do shealbhú na litearthachta sa Bhéarla, agus tá siad lárnach le ceadú dóibh siúd le fadhbanna litearthachta páirt iomlán a ghlacadh san oideachas agus sa tsaol ghairmiúil. Baintear úsáid fhorleathan as áiseanna don deachtú le

haithint cainte chun téacs scríofa a chumadh ó bhéal. Is mór an cuidiú áiseanna mar litreoirí agus gramadóirí. Ceadaíonn an tsintéis téacs a léamh os ard. Tacaíonn an tsintéis fosta le sealbhú na litearthachta, mar go léirítear an nasc idir litreacha agus fuaimeanna. Tá sé úsáideach chun téacs a cheartú ó chluais.

Tá tábhacht ar leith ag baint le taighde teangeolaíochta (Cuid I) d'áiseanna agus d'acmhainní a thacaíonn le sealbhú na litearthachta, m.sh. don scagthástáil, do mheasúnú agus d'áiseanna tacaíochta foghlama. Caithfidh siad seo a bheith bunaithe ar an méid seo:

(i) saineolas ar struchtúir fuaime agus ar chóras scríofa na Gaeilge – na gnéithe struchtúracha ar a mbraitheann chomh héasca agus chomh tapa is a dhéantar sealbhú ar an litearthacht (§2.7.3). Tá córas saibhir consan sa Ghaeilge atá an-éagsúil ó chóras an Bhéarla (§2.4). Mar an chéanna do chóras fónaice na Gaeilge: tá an nasc idir litir agus fuaim (córas fónaice) an-difriúil ó chóras an Bhéarla. Cé go bhfuil córas na Gaeilge doiléir, tá sé rialta (i gcomparáid le córas an Bhéarla) agus oiriúnach do mhodh teagaisc fónaice. Ach is é an gnáthchleachtas ná meabhrú a dhéanamh ar fhocail iomlána [14], modh nach n-oireann don teanga (féach §2.7.3) agus atá deacair d'fhoghlaimeoirí – go háirithe d'fhoghlaimeoirí le disléicse. Tá idirdhealú na gconsan mar ghné lárnach sa chóras fónaice: mar sin, do theagasc na fónaice, is gá dúshraith d'fheasacht (fóineolaíochta) ar idirdhealú na bhfuaimeanna.

(ii) tuiscint ar shealbhú na litearthachta Gaeilge, agus tionchar an chomhthéacs dhátheangaigh.

15.4.2 Cad é atá ar fáil go dtí seo?

Áiseanna don deachtú: níl siad seo ar fáil, ach tá forbairt á déanamh ar an aithint cainte (Caibidil 9).

Litreoirí agus gramadóirí: tá siad in úsáid go forleathan agus ar díol ag Cruinneog .

Áiseanna sintéiseacha: tá córais shintéiseacha ar fáil saor in aisce, mar aon le léitheoir ar líne agus áiseanna don oideachas, s.rl. (féach thíos agus Caibidil 8).

Uirlisí agus acmhainní don mheasúnú: níl a léithéid ann don Ghaeilge. Is don Bhéarla na hacmhainní don scagthástáil, don mheasúnú – rud atá iomlán mí-oiriúnach don Ghaeilge [15]. Fágann seo na síceolaithe oideachais a dhéanann measúnú ar pháistí le Gaeilge, agus múinteoirí tacaíochta foghlama gan na huirlisí atá riachtanach dóibh [16].

Taighde ar shealbhú na litearthachta Gaeilge: Tá taighde tosaigh déanta agus sonraí bailithe i scoileanna Gaeltachta agus i nGaelscoileanna [17-21], ach tá i bhfad níos mó de dhíth.

Áiseanna tacaíochta don litearthacht: le treoir ón taighde thuasluaite, tá forbairt ar siúl ag foireann ABAIR i gColáiste na Tríonóide ar áiseanna do shealbhú na litearthachta Gaeilge, a bhaineann úsáid as glórtha agus acmhainní na sintéise (Caibidil 14). Cé gur dírithe ar an ghnáthfhoghlaimeoir atá na háiseanna seo, chuideodh siad le daltaí le disléicse, agus tá sé mar sprioc go mbainfear úsáid astu mar bhunús d'áiseanna tacaíochta teanga amach anseo. D'fhoghlaimeoirí óga, díríonn an t-ardán idirghníomhach Lón don Leon [22] ar an fheasacht fóineolaíochta agus ar bhunscileanna litearthachta, le hamhráin, scéalta, pictiúir agus gníomhachtaí a chumadh chun struchtúr fuaime agus fónaice na Gaeilge a léiriú. D'fhoghlaimeoirí níos sine atá in ann an méarchlár a úsáid, díríonn an t-ardán foghlama An Scéalaí [23] ar scríobh agus léamh na Gaeilge, maraon le scileanna labhartha agus cluasthuisceana. Léann an glór sintéiseach téacs na ndaltaí amach, rud a chuidíonn an fhoirm scríofa agus labhartha a nascadh, agus moltar do dhaltaí a gcuid ábhar scríofa a cheartú ó chluais. Tá litreoir, gramadóir agus foclóir ceangailte leis an áis, agus tugtar nodanna faoi bhotúin don fhoghlaimeoir i bhfoirm scríofa agus labhartha (an tsintéis). Bailítear sonraí na ndaltaí ar mhaithe le taighde agus le forbairt na háise (féach Caibidil 14).

15.4.3 Riachtanais don Todhchaí

Áiseanna don deachtú a aithníonn cainteoirí agus canúintí éagsúla, agus, go háirithe, glórtha páistí.
Áiseanna sintéise: le glórtha agus canúintí éagsúla agus le glórtha páistí.

Litreoirí agus gramadóirí: beidh uasdátú de dhíth de réir mar a thagann an Plean Digiteach ar aghaidh.

Taighde ar shealbhú na litearthachta agus an disléicse: tá taighde forleathan ar shealbhú na litearthachta i scoileanna Gaeltachta, i nGaelscoileanna agus i scoileanna Béarla de dhíth (i) chun sonraí a bhailiú don scagthástáil agus don mheasúnú agus (ii) mar bhunús d'áiseanna tacaíochta d'fhoghlaim na litearthachta. Beidh an taighde seo chun tairbhe gach foghlaimeoir.
Uirlisí ar líne don scagthástáil agus don mheasúnú: tá sraith uirlísí ar line de dhíth, bunaithe sa taighde agus ag freagairt do struchtúir fuaime agus fónaice na Gaeilge. De ghnáth, déantar measúnú ar an fheasacht fóineolaíochta (idirdhealú agus láimhseáil ar fhuaimeanna na teanga) mar chuid de scagthástáil litearthachta. Tá ábhair ar líne le fuaimeanna dúchasacha an-oiriúnach, mar nach mbeidh cruinneas na measúnuithe ag brath ar urlabhra dúchasach a bheith ag an té atá i mbun na tástála. Beidh comhoibriú idir teangeolaithe agus síceolaithe oideachais de dhíth don obair seo.

Áiseanna tacaíochta foghlama don litearthacht: tá géarghá le háiseanna tacaíochta foghlama do dhaltaí le fadhbanna litearthachta agus a gcuid múinteoirí – bunaithe ar thaighde leanúnach sa réimse. Nuair a nasctar teicneolaíochtaí agus acmhainní éagsúla (an tsintéis, aithint cainte, litreoirí, gramadóirí, foclóirí, s.rl.) beidh féidearthachtaí ann d'áiseanna cumhachtacha le caint dhúchasach do na canúintí éagsúla. Moltar dlúth-chomhoibriú leis an fhorbairt leathan ar áiseanna oideachasúla (Caibidil 14).

Oiliúint ar speisialtóirí: deiseanna oiliúna sna réimsí cuí a chur ar fáil do lucht gairme – síceolaithe oideachais agus múinteoirí tacaíochta foghlama a bhfuil cúram daltaí le Gaeilge orthu (féach §15.6).

15.5 Áiseanna agus acmhainní don teiripe cainte

15.5.1 Cén tábhacht atá leis?

Bíonn teiripeoirí cainte ag plé le gach uile shaghas deacracht sa chumarsáid labhartha agus is minic gurb iadsan an nasc idir na taighdeoirí a dhéanann forbairt ar acmhainní agus áiseanna agus lucht úsáide na n-áiseanna. Tá baint ag gach gné den Phlean Digiteach leis an obair a bhíonn ar siúl ag lucht teiripe cainte do Ghaeilgeoirí, agus tá ionchur uathu i bpróiseas na forbartha lárnach chun gur féidir iomlán tairbhe a bhaint as.

Is riachtanas bunúsach do theiripe Gaeilge uirlisí agus treoirlínte a bheith ar fáil don mheasúnú atá bunaithe ar shonraí normatacha don teanga, chomh maith le cláir don idirghabháil teiripe dírithe ar fhadhbanna ar leith, atá deartha do chórais na teanga. Agus is riachtanas níos bunúsaí fós go mbíonn saineolas ag teiripeoirí ar struchtúr agus ar shealbhú na teanga, agus tuiscint ar thionchar an chomhthéacs dhátheangaigh.

15.5.2 Cad é atá ar fáil go dtí seo?

Is trí Bhéarla amháin a chuirtear oiliúint ar theiripeoirí in Éirinn, agus ní dhéantar teagasc ar theangeolaíocht na Gaeilge, ar shealbhú na teanga ná ar réimsí eile atá tábhachtach do na teiripeoirí atá le freastal ar Ghaeilgeoirí.

Tá taighde déanta i gColáiste na hOllscoile Corcaigh agus in Ollscoil na Gaillimhe, a dhíríonn go háirithe ar mheasúnú, agus ar uirlisí measúnaithe don Bhéarla a chur in oiriúint don Ghaeilge [24-29]. Tá grúpa teiripeoirí le Gaeilge, Neart, lonnaithe i nGaillimh ar suim leo sealbhú na teanga sa chomhthéacs dhátheangach. Ach ar an iomlán, tá bearna mhór sa réimse, agus easpa acmhainní do theiripeoirí.

Ní bhaintear úsáid as áiseanna teicneolaíochta don tástáil, don mheasúnú ná don idirghabháil teiripe don Ghaeilge. Níl áiseanna cúnta cumarsáide ann ach an oiread, seachas an léitheoir scáileáin atá luaite thuas, dóibh siúd faoi mhíchumas radhairc.

15.5.3 Riachtanais don Todhchaí

Cuirfidh an buntaighde teangeolaíochta agus an fhorbairt ar acmhainní i gCuid I lenár dtuiscint ar struchtúr na teanga mar a labhraítear inniu í agus cuirfidh siad sonraí normatacha ar fáil ar shealbhú na teanga agus na litearthachta sa chomhthéacs dátheangach – mar bhunús d'uirlisí measúnaithe don Ghaeilge.

Tá impleachtaí ag na bunteicneolaíochtaí agus na háiseanna i gCuid II agus III den Phlean do mhórán gnéithe teiripe – do theicneolaíochtaí cúnta mar AAC, d'uirlisí measúnaithe agus d'áiseanna don idirghabháil teiripe. I gcás mórán áiseanna a thógtar d'fhoghlaimeoirí Gaeilge, d'fhéadfaí leaganacha a fhorbairt chun tacú le hidirghabháil teiripe d'fhadhbanna urlabhra agus teanga. Mar shampla, bheadh samhlacha de phatrúin urlabhraíochta agus prosóide na Gaeilge – lena gceanglaítear aiseolas físiúil don úsáideoir – thar a bheith úsáideach sa teiripe. Mar an chéanna, d'fhéadfaí an áis oiliúna d'fheasacht fóineolaíochta agus do bhunscileanna litearthachta (§15.4) a chur in oiriúint don idirghabháil teiripe.

Chun go mbeidh aschuir acmhainní agus áiseanna an Phlean Dhigitigh chun leasa na teiripe cainte, moltar cúrsa oiliúna a chur ar fáil do speisialtóirí teiripe a bheidh ag obair le Gaeilgeoirí. Tá a leithéid de dhíth chun tacú leo ina ngairm agus chun muinín a spreagadh iontu nuair atá siad ag plé le Gaeilgeoirí [30]. Tá gá lena leithéid de speisialtóirí teiripe Gaeilge mar chomhpháirtithe i ngach gné taighde sa réimse, sa bhuntaighde ar shealbhú na teanga agus na litearthachta, sa dearadh ar uirlísí don mheasúnú; sa dearadh ar áiseanna don idirghabháil teiripe, do thástáil agus do scaipeadh na n-acmhainní agus n-áiseanna.

15.6 Moltaí Ginearálta

1. Taighde agus bailiúchán ar shonraí normatacha mar bhunús don fhorbairt ar uirlisí measúnaithe agus ar áiseanna tacaíochta foghlama agus ar áiseanna idirghabhála teiripe.
2. Áiseanna, uirlisí agus acmhainní a fhorbairt atá deartha don Ghaeilge agus fréamhaithe sa taighde, m.sh. gléasanna cumarsáide cúnta (AAC); comhéadain (m.sh do léitheoirí scáileáin); uirlisí ar

líne don scagthástáil agus don mheasúnú ar chumas urlabhra, teanga agus litearthachta; áiseanna tacaíochta foghlama agus áiseanna idirghabhála don teiripe.

3. 'Dearadh do Chách' [6] a bheith mar bhunphrionsabal i ngach gné forbartha.
   a. Dearadh ar áiseanna chun go mbeidh siad, nuair is féidir, inúsáidte ag daoine faoi mhíchumas.
   b. Bunteicneolaíochtaí a thógáil atá oiriúnach do gach úsáideoir, m.sh. córais shintéiseacha agus aitheanta cainte oiriúnach do chainteoirí agus canúintí éagsúla, agus go háirithe, do pháistí.
   c. Moltar go mbeadh ráiteas le gach iarratas ar mhaoiniú, a mhíníonn na himpleachtaí dóibh siúd faoi mhíchumas.
4. Líonraí Míchumais a bhunú a nascann grúpaí taighde agus daoine faoi mhíchumas, a mholann tosaíochtaí na forbartha, a ghlacann páirt san fhorbairt. Moltar líonraí mar:
   a. Líonra Comhairleach ar Mhíchumas, a chuireann comhairle ar ghrúpaí taighde faoi na riachtanais is práinní, agus a scaipeann eolas i measc daoine faoi mhíchumas ar fhéidearthachtaí an Phlean. Bheadh ballraíocht leathan de dhíth, ina measc, teiripeoirí cainte, síceolaithe oideachais, múinteoirí tacaíochta foghlama, eagrais a bhaineann le míchumas, Ranna Stáit, daoine faoi mhíchumas agus iad siúd a thacaíonn leo.
   b. Líonraí do mhíchumas ar leith, i gcás áis dírithe ar mhíchumas ar leith, tá gá le líonra – le húsáideoirí, múinteoirí, speisialtóirí – le bheith páirteach i ndearadh, measúnú agus scaipeadh na háise.
   c. Grúpa speisialtóirí gairme don Ghaeilge – Teiripeoirí Cainte/Síceolaithe Oideachais/ Múinteoirí Tacaíochta: moltar líonraí oifigiúla a bhunú ar speisialtóirí gairme don Ghaeilge, a bheidh ag freastal ar an tír ar fad, na Gaeltachtaí uile chomh maith le teaghlaigh agus foghlaimeoirí taobh amuigh den Ghaeltacht.
5. Oiliúint ar Speisialtóirí Gairme: tá géarghá le cúrsaí don fhorbairt ghairmiúil a thacóidh leis na speisialtóirí gairme atá ag plé le míchumas urlabhra, teanga, litearthachta agus cumarsáide. Tá oiliúint de dhíth a chlúdaíonn réimsí mar struchtúir teanga agus fónaice na Gaeilge; sealbhú teanga, cainte agus litearthachta agus tionchar an chomhthéacs dhátheangaigh laistigh agus lasmuigh den Ghaeltacht; buntaighde ar an dátheangachas agus ar na buntáistí d'fhorbairt an linbh; an disléicse; oiliúint ar úsáid na n-áiseanna agus na n-acmhainní a thagann chun tosaigh ón Phlean Digiteach. Tá géarghá leis na speisialtóirí oilte seo mar chomhpháirtithe san fhorbairt.
6. Líne Chabhrach Phoiblí le heolas, comhairle agus tacaíocht ó speisialtóirí: faoi láthair is minic míthuiscint taobh thiar den chomhairle a fhaigheann tuismitheoirí agus cúramóirí – rud atá iomlán míshásúil – go háirithe, nuair a mholtar an teanga baile a athrú, gan an Ghaeilge a theagasc don pháiste, s.rl. Moltar líne chabhrach lárnach a chur ar fáil, inár féidir leo siúd faoi mhíchumas, a dtuismitheoirí agus lucht cúraim comhairle agus tacaíocht a fháil ó shaineolaithe gairme agus ó thaighdeoirí acadúla sna réimsí cuí.

Tagairtí

[1] P. B. Denes, P.B., and Pinson, E. N., (1993) The Speech Chain: The Physics and Biology of Spoken Language, W. H. Freeman & Co Ltd.

[2] U.S. Centers for Disease Control and Prevention. (2018), https://www.disabled-world.com/disability/statistics/1in4.php

[3] Disability and unemployment: at https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Disability_statistics_-_need_for_assistance#Disabilities_and_the_labour_market]

[4] EBU:Euroblind information at: http://www.euroblind.org/about-blindness-and-partial-sight/facts-and-figures

[5] National Council for the Blind, Ireland (NCBI): https://www.ncbi.ie/facts-about-sightloss/

[6] Stephanidis, C., (Ed.) (2001). User Interfaces for All: Concepts, Methods and Tools. Lawrence Erlbaum Associates.

[7] Feifei, X., Barker, J., and Christensen, H., (2019), Phonetic Analysis of Dysarthric Speech Tempo and Applications to Robust Personalised Dysarthric Speech Recognition, in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing

[8] Yamagishi, J., Veaux, C., King, R., & Renals, S. (2012) Junichi Yamagishi, Christophe Veaux, Simon King, and Steve Renals. Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction. Acoustical Science and Technology, 33(1):1-5

[9] Guenther, F. H. (2015). The neural control of speech: From computational modeling to neural prosthesis, Proceedings of Int. Congress of Phonetic Sciences, Glasgow, Scotland.

[10] McGuirk, R. (2015). Exploration of the Use of

Irish Language Synthesis with a Screen Reader in the Teaching of Irish to Pupils with Vision Impairment, MPhil Dissertation, School of Linguistic, Speech and Communication Sciences, Trinity College, Dublin, Ireland.

[11] Daisybooks information at: http://www.daisy.org/daisypedia/daisy-digital-talking-book

[12] Dyslexia Association of Ireland: https://www.dyslexia.ie/

[13] Dep of Education & Science: http://www.education.ie/en/Publications/Statistics/Key-Statistics/),

[14] Stenson, N., & Hickey, T. (2018). Understanding Irish Spelling. An Chomhairle um Oideachas Gaeltachta agus Gaelscolaíochta.

[15] Nic Aindriú, S., Duibhir, P. Ó., & Travers, J. (2021). A Survey of Assessment and Additional Teaching Support in Irish Immersion Education. Languages, 6(2), 62.

[16] Barnes, E. (2017) Dyslexia Assessment and Reading Interventions for Pupils in Irish- Medium Education: Insights into current practice and considerations for improvement, M.Phil. Dissertation, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Ireland

[17] Barnes, E., Ní Chasaide, A. and Ní Chiaráin, N. (2018) The design and pre-testing of literacy and cognitive tasks in Irish and English, in Proceedings of Literacy Association of Ireland 42nd International Conference, Dublin

[18] Barnes, E., Ní Chasaide, A., Ní Chiaráin, N. (2021) Bilingual phonological awareness: when interdependence becomes interference, in Proceedings of the 15th Congress of the International Association for the Study of Child Language.

[19] Parsons, C. E., & Lyddy, F. (2016). A longitudinal study of early reading development in two languages: comparing literacy outcomes in Irish immersion, English medium and Gaeltacht schools. International Journal of Bilingual Education and Bilingualism, 19(5), 511-529.

[20] Murphy, D., & Travers, J. (2012). Including young bilingual learners in the assessment process: A study of appropriate early literacy assessment utilising both languages of children in a Gaelscoil. Special and inclusive education: A research perspective, 167-185.

[21] Barnes, E. (2021) Predicting dual-language literacy attainment in Irish-English bilinguals: language-specific and language-universal contributions, PhD Thesis, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Ireland

[22] Ní Chasaide, A., Ní Chiaráin, N., Berthelsen, H., Wendler, C., Murphy, A., Barnes, E. and Gobl, C. (2019). Leveraging phonetic and speech research for Irish language revitalisation and maintenance. Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, pp. 994-998.

[23] Ní Chiaráin, N., Ní Chasaide, A., (2019) An Scéalaí: autonomous learners harnessing speech and language technologies, SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education, Graz, Austria

[24] O' Malley, M.P. & Antonijevic, S. (2020). Adapting MAIN to Irish (Gaeilge). ZAS Papers in Linguistics 64: 127-138.

[25] O'Toole, C, Fletcher, P (2012) Profiling vocabulary acquisition in Irish, in Journal of Child Language, 39

[26] O'Toole, C. & Fletcher, P. (2008) Developing assessment tools for bilingual and minority language acquisition, Journal of Clinical Speech and Language Studies, 16 :12-27

[27] O'Toole, C (2013) Using parent report to assess bilingual vocabulary acquisition: A model from Irish, in Gathercole, VCM (eds), Solutions for the Assessment of Bilinguals. Bristol: Multilingual Matters

[28] O'Toole, C,Hickey, TM (2013) Diagnosing language impairment in bilinguals: Professional experience and perception, in Child Language Teaching & Therapy, 29 :91-109

[29] Antonijevic, S., Durham, R., & Ní Chonghaile, Í. (2017). Language performance of sequential bilinguals on an Irish and English sentence repetition task. Linguistic Approaches to Bilingualism, 7, 3-4, 359-393.

[30] Harnett, S. & O'Toole, C. (2010) 'Speech and language therapists training and confidence, Journal of Clinical Speech and Language Studies, 17 :57-73

[31] Barnes, E., Morrin, O., Ní Chasaide, A., Cummins, J., Berthelsen, H., Murphy, A., Nic Corcráin, M., O' Neill, C., Gobl, C., Ní Chiaráin, N. (2022), AAC don Ghaeilge: Prototype Development of Speech-Generating Assistive Technology for Irish, Proceedings of the CLTW 4 @ LREC2022 , pages 127–132

Caibidil 16
Conclúid: an Plean a Chur i bhFeidhm

16.1 An Plean Digiteach: próiseas fadtéarmach, le taighde comhtháite

Cé go bhfuil an Plean Digiteach curtha i láthair sa cháipéis seo i bhfoirm caibidlí/réimsí aonaracha, níor cheart breathnú air mar liosta de theicneolaíochtaí atá le forbairt, ach mar chlár leathan ina mbeidh na gnéithe éagsúla taighde ag tacú lena chéile. De réir mar a thiocfaidh teicneolaíochtaí agus áiseanna chun tosaigh, nascfar le chéile iad ar bhealaí nár samhlaíodh go fóill. Is nuair a fhífear le chéile iad a bhraithfear neart agus tionchar forleathan na teicneolaíochta.

Is plean oibre fadtéarmach oscailte é an Plean Digiteach agus tá i bhfad níos mó i gceist ná áiseanna faoi leith a sholáthar. Is bunú dúshraithe faoi choinne earnáil dhúchasach theicneolaíocht na Gaeilge an mhórsprioc, earnáil atá dlúthnasctha le pobal na Gaeilge. Cuirtear an bhéim sa chaibidil seo agus sa Réamhrá (Caibidil 1) ar na gníomhaíochtaí agus ar na modhanna oibre a bhfuil gá leo chun go dtiocfaidh tairbhe iomlán as torthaí an Phlean do phobal na Gaeilge agus do chaomhnú na teanga.

Sa chéad áit, cuirtear an bhéim ar thógáil an bhonneagair sa tír seo agus ar oiliúint na dtaighdeoirí, a cheadóidh forbairt leanúnach ar theicneolaíochtaí urlabhra agus teanga, atá bunaithe i saineolas ar struchtúr na Gaeilge, cónasctha le pobal na Gaeilge agus ag freagairt dá riachtanais. Sa dara háit, chun slabhra iomlán forbartha a chur i gcrích, ó thaighde bunteicneolaíocht áiseanna scaipeadh i measc an phobail beidh foirne idirdhisciplíneacha de dhíth ina bhfuil sruthanna comhthreomhara taighde agus forbartha, agus taighdeoirí as sruthanna éagsúla ag obair as lámha a chéile. Sa tríú háit cuirtear béim ar an ngá le rannpháirtíocht an phobail i ngach gné den Phlean.

16.2 Cé dó é?

Is dúinn ar fad an Plean Digiteach. Beidh sé ag freastal orthu siúd ar fad ar suim leo an teanga – cainteoirí dúchais na Gaeltachta, Gaeilgeoirí taobh amuigh den Ghaeltacht, daltaí scoile, a gcuid múinteoirí, na tuismitheoirí a chuidíonn leo sa bhaile, foghlaimeoirí ag gach leibhéal in Éirinn agus i gcéin, inimircigh – daoine gach áit ar suim leo an Ghaeilge. Tá béim sa Phlean orthu siúd faoi mhíchumas atá, de cheal áiseanna cuí, go mór faoi mhíbhuntáiste i saol na Gaeilge.

16.3 Tionchar an Phlean

Le breacadh na ré digití don Ghaeilge, beidh cumhacht na dteicneolaíochtaí nuálacha ag tacú le cothú agus caomhnú phobal na Gaeilge, agus ag éascú naisc chumarsáide idir pobail atá scartha amach óna chéile. Cumasófar úsáid níos forleithne na teanga, nuair a bheidh sí ar fáil inár gcuid áiseanna digiteacha. Beidh an fhéidearthacht ag aos óg digiteach na linne seo an teanga a úsáid sa ghnáthshaol sóisialta laethúil – ag cuidiú leis an bhearna a shárú idir a n-oidhreacht teanga agus saol an lae inniu, ina bhfuil an teicneolaíocht fite fuaite.

Leis an infheistíocht cheart, tá féidearthachtaí móra ann do theagasc agus d'fhoghlaim na Gaeilge, lena dtacaíonn gnéithe éagsúla den Phlean. De thairbhe na mbunchóras teicneolaíochta beifear in ann ardáin nuálacha 'chliste' teagaisc a fhorbairt, m.sh., cluichí idirghníomhacha ilmhódacha lena ngabhann carachtair fhíorúla. Tabharfaidh a leithéid seo d'áis deis d'fhoghlaimeoirí scileanna cumarsáide a chothú ar bhealach spraíúil, agus cuirfear lena n-inspreagadh agus iad ag caint le cainteoirí dúchais (fíorúla). Beidh taighde teangeolaíochta agus na hacmhainní as Cuid I den Phlean á n-úsáid don ábhar teagaisc. Beidh an taighde eolaíoch ar shealbhú na teanga agus ar chéimeanna foghlama na ndaltaí – ó thosaitheoirí go húsáideoirí cumasacha – ag treorú an churaclaim agus mar bhunús d'áiseanna nua chun measúnú agus ábhair tacaíochta foghlama a chur ar fáil, agus dá thoradh sin beidh múinteoirí in ann díriú mar is cóir orthu siúd a bhfuil deacrachtaí acu.

Ar bhealaí éagsúla, trí chumhacht na teicneolaíochta a chur i lámha phobal labhartha na Gaeilge, trí eispéireas níos dearfaí d'fhoghlaimeoirí laistigh agus lasmuigh den Ghaeltacht, beidh an Plean Digiteach ag tacú le haidhm na Straitéise 20 Bliain don Ghaeilge [1], is é sin an dátheangachas a thabhairt i gcrích in Éirinn.

16.4 Comhpháirtíocht le Pobal na Gaeilge

Braithfidh éifeacht agus tionchar an Phlean go hiomlán ar an pháirt a ghlacann pobal na Gaeilge ann. I bhforbairt na teicneolaíochta, ní hionann don Ghaeilge agus do mhórtheangacha mar an Béarla, ina bhfuil an brabús mar phríomhaidhm ag na gnólachtaí teicneolaíochta, agus pobal ollmhór ag ceannach a gcuid earraí. I gcás na Gaeilge agus i gcás teangacha atá i mbaol, caithfidh tosaíochtaí an taighde agus roghnú na n-áiseanna atá le forbairt a bheith bunaithe ó thús ar riachtanais an phobail, agus an bhéim ar na háiseanna is mó a bhfuil gá leo. Mar sin, cé go bhfuil ionaid taighde i gceannas na hoibre, beidh comhpháirtíocht leis an phobal, agus iad ina rannpháirtithe gníomhacha san fhorbairt, agus le lucht úsáide na n áiseanna lárnach i ngach gné, m.sh.:

- Treoir ar thosaíochtaí taighde: le freagairt do na riachtanais is práinní.
- Dearadh na n-áiseanna: tá ionchur ó na húsáideoirí de dhíth ón tús ionas go mbeidh na feidhmchláir éifeachtach agus ábhartha le haghaidh na n-úsáideoirí.
- Comhpháirtíocht i ngnéithe forbartha: i mórán réimsí forbartha, tá ionchur de dhíth ón phobal, ó ghrúpaí agus ó dhaoine aonraic (féach plé ar shluafhoinsiú agus ar líonraí, §16.9).
- Tástáil: tá tástáil leanúnach ó úsáideoirí de dhíth lena chinntiú go bhfuil na teicneolaíochtaí agus na háiseanna seasmhach agus iomlán oiriúnach. Le measúnú ar áis, is féidir na bunteicneolaíochtaí atá leabaithe ann a mheasúnú chomh maith (m.sh., measúnú ar ghuthanna sintéiseacha mar ghné den tástáil ar chluiche oideachasúil [2]).
- Forleathnú: tá caidreamh leanúnach le lucht labhartha na Gaeilge de dhíth chun go mbaintear úsáid fhorleathan as na teicneolaíochtaí.

Tá moltaí sna míreanna a leanas ar bhealaí ar féidir comhpháirtíocht le pobal na Gaeilge a chothú.

16.5 Ionaid bairr feabhais taighde: an bonneagar taighde

Ionaid bairr feabhais taighde:

- Foirne idirdhisciplíneacha: tá scileanna teicniúla (innealtóireachta agus ríomheolaíochta) de dhíth chomh maith le hardchumas teanga agus teangeolaíochta sa Ghaeilge. Beidh na hionaid/réimsí atá ag gníomhú faoi láthair le neartú agus scála agus réimsí an taighde le leathnú amach.
- Maoiniú leanúnach: tá seo ríthábhachtach chun forbairt fhadtéarmach a chinntiú.
- Aistriú na n-aschur taighde > áiseanna don phobal: chun é seo a chur i gcrích, beidh forbairt áiseanna le déanamh ag foirne taighde ag an am chéanna le buntaighde. Beidh tacaíocht theicniúil leanúnach de dhíth, mar a phléitear in §16.8 thíos.

16.6 Oiliúint ar thaighdeoirí: an bonneagar daonna

Don fhorbairt fhadtéarmach, is é an rud is tábhachtaí ar fad ná an bonneagar daonna – na taighdeoirí a chuirfidh an Plean i gcrích. Go dtí seo, ba é an chonstaic ba mhó ag grúpaí taighde sa réimse ná easpa taighdeoirí le hardchumas Gaeilge mar aon le scileanna teicniúla agus teangeolaíochta. I ngnáth-thionscadal (Béarla) den sórt seo, is féidir teacht go héasca ar thaighdeoirí sároilte, as tír ar bith ar domhan – ach ní hionann don Phlean Dhigiteach seo. Tá infheistiú de dhíth i gcúrsaí teagaisc idirdhisciplíneacha chun oiliúint a chur ar thaighdeoirí a bhfuil ardchaighdeán Gaeilge acu, tuiscint ar struchtúr na teanga, scileanna innealtóireachta urlabhra agus scileanna próiseála teanga. Don oiliúint ar thaighdeoirí na todhchaí, moltar go mbunófaí/go neartófaí:

- Cúrsaí idirdhisciplíneacha bunchéime agus iarchéime a fhreagraíonn do riachtanais an Phlean Dhigitigh. Tá cúrsaí i gColáiste na Tríonóide a dhéanann freastal go pointe áirithe ar na riachtanais idirdhisciplíneacha: bunchéim chomhtháite sa Ríomheolaíocht, Teangeolas agus Teanga (Gaeilge) agus cúrsa máistreachta Próiseáil Urlabhra agus Teanga. Is céimithe ó na cúrsaí seo mórán de na taighdeoirí atá ag saothrú sa ghort faoi láthair. Moltar na cúrsaí sin a neartú agus a leathnú amach chun freastal ar bhealach níos iomláine ar riachtanais theicniúla, teanga agus teangeolaíochta an Phlean (féach fosta moltaí in §16.7 i dtaca le féidearthachtaí i dtaobh cúrsaí a pháirtlonnú sna ceantair Ghaeltachta).
- Scéim sparánachtaí a bhunú a mheallfadh cainteoirí dúchais agus cainteoirí líofa Gaeilge le tabhairt faoina leithéid de chúrsaí.
- Maoiniú dochtúireachta agus iardhochtúireachta do thaighdeoirí atá ag obair le grúpaí taighde atá dírithe ar an Phlean Dhigiteach – tá sé tábhachtach le deiseanna a chur ar fáil do chéimithe oilte leanúint den obair san earnáil, agus a bpáirt a ghlacadh mar cheannairí taighde don todhchaí. Sa chás go bhfuil gá le speisialtóirí teicneolaíochta nach bhfuil Gaeilge acu, moltar an deis a thabhairt dóibh í a fhoghlaim agus tuiscint a fháil ar struchtúr na teanga.
- Oiliúint san fhiontraíocht, mar pháirt den oiliúint ar scoláirí ag gach leibhéal. Tá a leithéid de chúrsaí ar fáil sna hinstitiúidí tríú leibhéal.

16.7 An Plean Digiteach agus an Ghaeltacht

Tá an Ghaeilge beo mar teanga phobail sa Ghaeltacht, ach go mór faoi bhrú. Moltar go ndíreofaí ar an nasc leis an Ghaeltacht agus ar na bealaí ar féidir leis an Phlean Dhigiteach tacú le cainteoirí dúchais na Gaeltachta. Mar nach bhfuil an Ghaeilge ar fáil i mórán áiseanna digiteacha, níl sí san áireamh i mórán gníomhaíochtaí ar líne ag aos óg digiteach na Gaeltachta. Chomh maith leis na pointí atá luaite in §16.4 – §16.6, moltar rannpháirtíocht le cainteoirí dúchais na Gaeltachta a chothú, ag coinneáil lena dtábhacht le haghaidh:

Múnlaí cainte: is iad na 'speisialtóirí cainte' a bhfuil gá leo do na corpais labhartha fhairsinge atá de dhíth i ngach canúint don taighde teangeolaíochta agus do theicneolaíocht na hurlabhra. Beidh taifeadadh ar láthair agus sluafhoinsiú dírithe ar na ceantair Ghaeltachta de dhíth (§16.9). Taighdeoirí na todhchaí: táthar ag súil go mbeidh cainteoirí dúchais na Gaeltachta lánpháirteach i bhforbairt an Phlean Dhigitigh. Leis an sprioc sin a chur chun cinn, moltar:

- An óige a mhealladh trí imeachtaí ina ndírítear go háirithe ar aos óg na Gaeltachta, m.sh. cuairteanna ó lucht taighde ar scoileanna Gaeltachta; ceardlanna do pháistí agus do dhéagóirí inar féidir leo spraoi leis na háiseanna; cuireadh chuig shaotharlanna taighde do dhaltaí na hidirbhliana – a léiríonn don aos óg na réimsí oiliúna tríú leibhéal agus féidearthachtaí gairme i bhforbairt teicneolaíochta na Gaeilge.
- Mic léinn Ghaeltachta a earcú do chúrsaí idirdhisciplíneacha céime agus iarchéime, mar atá luaite thuas in §16.6. Le sparánachtaí, bheadh na cúrsaí indéanta agus mealltach.
- Páirt-teagasc sa Ghaeltacht (ollscoil ¬¬– Gaeltacht) do chúrsaí oiliúna taighdeoirí: mar atá luaite thuas in §16.6. B'fhiú a fhiosrú an bhféadfaí páirteanna de na cúrsaí teagaisc a lonnú sna ceantair Ghaeltachta – mar shampla le teagasc ar líne nuair a fhóireann sé sin. Bheadh pleanáil chúramach de dhíth.

Fiontraíocht le lonnú ollscoile & Gaeltachta: san earnáil seo, beidh féidearthachtaí amach anseo do ghnólachtaí a dhéanann margaíocht ar áiseanna faoi leith (féach §16.10). Is suíomh nádúrtha an Ghaeltacht do thionscadal teanga, agus chuirfeadh sé le deiseanna fostaíochta ar fáil do chéimithe oilte ina gceantar dúchais. I gcás gnólacht campais, a mbeadh baint aige le hionad taighde (ionad ollscoile nó eile) b'fhiú lonnú comhroinnte a fhiosrú, le gníomhaíochtaí gnó i mol digiteach Gaeltachta agus tacaíocht leanúnach ón ionad taighde i dtaca le forbairtí nuálacha teicneolaíochta. Is ceist don tréimhse fhadtéarmach í seo, ach is fiú an phleanáil a dhéanamh ón tús.

Tá comhlonnú cúrsaí oiliúna agus fiontraíochta tábhachtach chun nach mbíonn na deiseanna oiliúna agus fostaíochta d'aos óg na Gaeltachta á sú amach as a gceantair dhúchais, ag cur le bánú na Gaeltachta agus le laghdú ar phobal na gcainteoirí dúchais atá faoi bhrú cheana féin. Is ceann de na buntáistí a ghabhann leis an ré dhigiteach go bhfuil sé indéanta fostaíocht shuimiúil dhúshlánach a bheith ar fáil ag aos óg na Gaeltachta ina n-áit dúchais féin, ag cur le todhchaí an phobail. Is fiú a lua nach gciallaíonn an bhéim seo ar chomhpháirtíocht Ghaeltachta nach mbeadh lántacaíocht fosta ann do thionscnaimh a bheidh lonnaithe taobh amuigh den Ghaeltacht. Táthar ag súil go dtiocfaidh méadú ar chomhpháirtíocht leathan le pobal labhartha na Gaeilge, cibé áit a bhfuil sé.

16.8 An bonneagar teicniúil: tacaíocht leanúnach do chórais

Mar go bhfuil slabhra iomlán forbartha i gceist – ó thaighde bunúsach go háiseanna atá dírithe ar ghrúpaí faoi leith – beidh tacaíocht theicniúil de dhíth ar dhá leibhéal:

Tacaíocht áitiúil: tá tacaíocht ríomhchlárúcháin de dhíth laistigh de na hionaid taighde, chun gur féidir córais sheasmhacha a thógáil as leaganacha píolótacha taighde. Tá gá fosta le comhéadain chun na córais teicneolaíochta a chur ag obair laistigh d'fheidhmchláir sheachtracha. Tá dífhabhtú agus tástáil chuimsitheach de dhíth ar na bogearraí ar fad. Beidh uasdátú leanúnach le déanamh, de réir mar a thiocfaidh athruithe sa teicneolaíocht, mar shampla i gcórais oibriúcháin na ríomhairí agus le gléasanna nua.

Tacaíocht lárnach: le himeacht aimsire, beidh gá le lárionad tacaíochta teicniúla le haschuir theicneolaíochta na n-ionad taighde éagsúil a scaipeadh i measc an phobail – i scoileanna, in oifigí, sa bhaile srl. – agus le déileáil le fadhbanna teicniúla a bhíonn ag úsáideoirí. Bheadh ról gníomhach ag a leithéid d'aonad i gcraobhscaoileadh eolais ar áiseanna nua atá á bhforbairt mar aon le háiseanna atá ar fáil cheana féin.

16.9 An nasc leis an phobal, le ranna stáit agus le grúpaí sainleasa

Chun tairbhe iomlán a bhaint as an Phlean Dhigiteach, tá caidreamh gníomhach leis an phobal, le líonraí agus grúpaí sainleasa agus le rannóga stáit ríthábhachtach. Bhraith an dul chun cinn go dtí seo go mór ar chomhoibriú lena leithéid de líonraí (m.sh. múinteoirí,

aistritheoirí, líonraí míchumais radhairc). I measc na ngrúpaí lena mbeidh comhoibriú ag teastáil, tá:

Líonraí Gaeilge: Beidh sé tábhachtach i bhforbairt na teicneolaíochta go mbeidh teagmháil ag na grúpaí taighde le heagrais agus grúpaí pobail atá bainteach leis an teanga, in Éirinn agus thar lear. Tá na meáin Ghaeilge i gceartlár shaol na Gaeilge: beidh na teicneolaíochtaí nuálacha chun tairbhe dóibh agus beidh ról lárnach acu i scaipeadh eolais ar áiseanna mar a thagann siad chun tosaigh. Tugann na meáin shóisialta bealaí eile le teagmháil a dhéanamh le daoine nach bhfuil ina mbaill de líonraí Gaeilge ach ar mhaith leo a bheith páirteach san fheachtas.

An Rialtas: Is plean náisiúnta an Plean Digiteach a bheidh ag tacú le gach roinn rialtais, agus atá tábhachtach do na ranna seo leanas go háirithe: an Roinn Oideachais; an Roinn Sláinte; an Roinn Leanaí, Comhionannais, Míchumais, Lánpháirtíochta agus Óige; agus an Roinn Breisoideachais agus Ardoideachais, Taighde, Nuálaíochta agus Eolaíochta. Cé go bhfuil an tionscnamh á bhunú ag an Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meán, beidh tacaíocht fhorleathan de dhíth ó ranna rialtais agus ó na páirtithe polaitíochta eagsúla. Beidh maoiniú suntasach de dhíth agus tá sé tábhachtach nach mbeidh de thoradh ar an bhuiséad don Phlean Dhigiteach go gcuirfear laghdú ar an bhuiséad atá ar fáil faoi láthair d'eagrais agus do phobail Ghaeltachta atá ag obair chun an teanga a chaomhnú.

Tugann an Plean Digiteach deis thábhachtach don Rialtas feoil a dhéanamh den bhriathar i dtaca le stádas bunreachtúil na Gaeilge mar an chéad teanga náisiúnta agus Achtanna na dTeangacha Oifigiúla, 2003 agus 2021 a chur i bhfeidhm [3]. Is céim chun tosaigh é don Straitéis 20 Bliain don Ghaeilge, 2010 – 2030 [4] agus do Phlean Gníomhaíochta 2018-2022 [4]. Tacóidh sé, mar shampla, le soláthar seirbhísí dátheangacha rialtais, agus cuideoidh na háiseanna cumarsáide labhartha le caidreamh labhartha nó scríofa ar líne le heagrais stáit a chur ar fáil trí Ghaeilge. Éascóidh sé fosta an fhorbairt ar áiseanna rochtana a gealladh dóibh siúd atá faoi mhíchumas san Acht um Míchumas, 2005 [5], a mhéad atá sé sin "indéanta agus oiriúnach".

Chomh maith le maoiniú agus tacaíocht don Phlean Dhigiteach, tá mórán bealaí praiticiúla ann ar féidir leis na ranna rialtais tacú le réimsí taighde faoi leith. Mar shampla, is féidir leo a chinntiú, nuair a dhéantar aistriúchán, go dtugtar na fotháirgí a bhaineann leis an phróiseas (na comhaid cuimhne aistriúcháin, na bunachair agus na gluaiseanna téarmaíochta) do na grúpaí taighde chun na córais aistriúcháin ríomhchuidithe a fheabhsú. Nuair a thógfar áis mar léitheoir ar líne don Ghaeilge, ba cheart go gcuirfí i bhfeidhm é i leathanaigh ghréasáin na ranna rialtais. Beidh idirghabháil agus comhordú le ranna rialtais ríthábhachtach i scaipeadh na n-áiseanna agus na n-acmhainní sa dóigh is go gcuirfear na háiseanna nuálacha ar fáil go forleathan do scoileanna, d'eagrais ghnó, d'eagraíochtaí stáit agus státurraithe, agus don ghnáthshaoránach.

Líonraí ar líne agus sluafhoinsiú: Beidh líonraí ar líne agus sluafhoinsiú ag cuidiú le gnéithe áirithe taighde. Le sluafhoinsiú is féidir tacaíocht an phobail a fháil le tabhairt faoi thascanna mar bhailiú nó anótáil sonraí, agus is cuidiú mór é seo nuair a bhíonn a lán obair láimhe fhadálach le déanamh. Mar sin féin, i gcás mionteanga mar an Ghaeilge, is 'slua' d'fhoghlaimeoirí ag leibhéil dhifriúla a bhíonn i gceist go minic, agus caitear na tascanna a roghnú le bheith in oiriúint don tslua agus bíonn maoirseacht de dhíth i dtaca le caighdeán na sonraí. Is féidir díriú ar 'shlua' faoi leith, nó monatóireacht a dhéanamh ar an ábhar a fhaightear. Ag brath ar na sonraí atá de dhíth is féidir modhanna éagsúla sluafhoinsithe a úsáid:

(i) Sluafhoinsiú oscailte: nuair nach bhfuil ardchaighdeán teanga ná saineolas de dhíth. Chuir sluafhoinsiú den tsaghas seo go mór leis na bunachair eolais ag Dúchas.ie [6] agus Meitheal Logainm.ie [7].
(ii) Sluafhoinsiú faoi mhonatóireacht: ina bhfuil grúpa saineolaithe a dhéanann eagarthóireacht ar an ábhar uile a thagann ón tslua, mar a rinneadh le logánú Gmail [8] agus WhatsApp [9].
(iii) Sluafhoinsiú atá dírithe ar ghrúpaí faoi leith: Mar shampla, agus corpais fhorleathana chainte MíleGlór á mbailiú don aithint cainte [10], déantar idirdhealú idir cainteoirí dúchais le canúintí éagsúla, Nua-Ghaeilgeoirí, foghlaimeoirí srl. I gcás bailiúchán ábhair do Vicipéid na Breatnaise, rinneadh roghnú roimh ré ar an 'slua' agus cuireadh oiliúint ar fáil don obair.

Líonraí oideachais: Tá béim sa Phlean Digiteach ar na himpleachtaí do theagasc na Gaeilge agus don oideachas trí Ghaeilge (ón naíonra go dtí an 3ú leibhéal) agus d'fhoghlaimeoirí lasmuigh den chóras oideachais in Éirinn agus i gcéin. Beidh ionchur ó na heagrais oideachais agus teagaisc de dhíth go leanúnach. I bhforbairt áiseanna, tá dlúth-chomhoibriú de dhíth idir taighdeoirí agus líonraí ina bhfuil foghlaimeoirí, múinteoirí, lucht deartha ábhair teagaisc srl ag cuidiú le dearadh na n-áiseanna, ag baint triail as leaganacha píolótacha, ag cuidiú le hábhar teagaisc den chéad scoth a chur ar fáil, ag cinntiú go mbaintear úsáid

fhorleathan as na háiseanna nuair a bhíonn siad réidh don phobal.

Líonraí aistritheoirí: Tá líonra mór d'aistritheoirí ag obair ar réimse leathan ábhar, idir litríocht agus doiciméid theicniúla, atá ag teacht ó ranna stáit agus go háirithe anois ó institiúidí an Aontais Eorpaigh. Tá ról lárnach ag na líonraí seo i dtástáil agus i measúnú na gcóras aistriúcháin ríomhchuidithe atá ag teacht chun tosaigh. Beidh buntáistí móra fiú le córais nach bhfuil iomlán críochnúil, sa mhéid is gur féidir leo próiseas an aistriúcháin a bhrostú agus a dhéanamh níos éifeachtaí agus fáil réitithe de thascanna leadránacha a bhíonn le déanamh arís agus arís eile. Déanfaidh na lárionaid atá ag gabháil don taighde seo eolas ar acmhainní a scaipeadh agus oiliúint a chur ar fáil ar úsáid na n ardán nuálach.

Líonraí míchumais agus rochtana: Beidh an Plean ag freastal orthu siúd atá faoi mhíchumas, agus atá go minic, de cheal easpa acmhainní agus áiseanna rochtana, gearrtha amach ó shaol na Gaeilge. Tá caidreamh le líonraí míchumais i bhforbairt áiseanna ríthábhachtach le cinntiú go bhfuil siad in oiriúint do na húsáideoirí agus go bhfuil teacht acu orthu. Bhí a leithéid de líonraí lárnach don fhorbairt a rinneadh ar bhreiseán Gaeilge don léitheoir scáileáin NVDA [11] (Caibidil 15). Is minic nach mbíonn eolas ag daoine faoi mhíchumas ar fhéidearthachtaí na dteicneolaíochtaí, nó ar a gcearta i dtaca leo. Mar sin, beidh comhoibriú le múinteoirí, lucht teiripe, le grúpaí míchumais agus daoine aonraic faoi mhíchumas lárnach chun freastal ar a gcuid riachtanas agus a gcás a chur chun tosaigh. Fearacht na réimsí eile forbartha, beidh na líonraí gníomhach i ndearadh agus i dtástáil na n-áiseanna, agus nuair a bheidh áis réidh le scaoileadh amach, beifear ag brath ar na líonraí céanna chun go mbainfear úsáid fhorleathan astu.

Líonraí taighde idirnáisiúnta: Tá athruithe ag teacht ar an teicneolaíocht ar luas lasrach, agus tá sé tábhachtach go mbeidh teicneolaíocht na Gaeilge chun tosaigh sa réimse. Chuige sin beidh teagmháil agus comhoibriú le saotharlanna agus le hionaid bairr feabhais i dtíortha eile riachtanach. I gcásanna áirithe tá acmhainní agus corpais á gcomhroinnt ag líonraí Eorpacha, rud atá tábhachtach i réimse mar an aistriúchán ríomhchuidithe, mar a bhfuil ELRC [12] ag cur feabhas ar an aistriúchán uathoibríoch i riaracháin phoiblí na hEorpa. Beidh leas fosta le baint as teagmháil le líonraí mar CLARIN [13] agus META-SHARE [14]. Tá tábhacht ar leith ag baint le líonra SiGUL [15], a chothaíonn teagmháil le comhghleacaithe atá ag obair ar theicneolaíochtaí do mhionteangacha eile.

16.10 Cúrsaí gnó

Mar gur ar mhaithe leis an Ghaeilge agus le pobal na Gaeilge an Plean Digiteach, agus mar go bhfuil sé á mhaoiniú ag an Stát, tá sé mar bhunphrionsabal go mbeidh na háiseanna a eascraíonn as á soláthar saor in aisce do phobal na Gaeilge, a mhéad is féidir. Mar sin féin, de réir mar a thiocfaidh borradh faoi na háiseanna teicneolaíochta, beidh deiseanna ann d'fhiontair ghnó agus do ghnólachtaí campais, a dhéanann margaíocht ar áiseanna faoi leith. Moladh i mír §16.7 go mbreathnófaí sa chás sin ar fhéidearthachtaí a leithéid de ghnólacht campais a bheith chomhlonnaithe idir an Ghaeltacht agus an t-ionad taighde as a n-eascraíonn an áis. Is ceist mheántéarmach agus fhadtéarmach í seo, ach is fiú pleanáil ina treo, m.sh. trí oiliúint fiontraíochta a chur ar mhic léinn agus ar thaighdeoirí; trí fhéidearthachtaí i dtaobh suíomhanna a phlé le heagrais mar Údarás na Gaeltachta; trí mhachnamh agus plé a dhéanamh ar pholasaí agus ar threoirlínte faoin sciar d'aon bhrabús a théann ar ais isteach i gclár taighde an aonaid as ar eascair sé.

Táthar ag súil go mbeidh tionchar ag an Phlean a théann níos faide ná pobal comhaimseartha na Gaeilge. Ba cheart go bhfeicfeadh gnólachtaí nach bhfuil bainteach leis an Ghaeilge gur fiú dóibh áiseanna Gaeilge a chur in úsáid ina gcuid ardán margaíochta. Tá go leor de na gnólachtaí móra idirnáisiúnta lonnaithe sa tír seo a bhfuil baint acu le teicneolaíocht teanga agus urlabhra. Mar chuid den fhreagracht shóisialta atá ar na gnólachtaí sin, moltar go n-iarrfadh an Rialtas orthu céatadán (beag) dá mbrabús a bhronnadh mar mhaoiniú don Phlean Dhigiteach le tacú le hearnáil teicneolaíochta Gaeilge a bhunú sa tír seo. Chomh maith le maoiniú, tá mórán bealaí ann ar féidir leis na gnólachtaí tacú leis an Ghaeilge, mar shampla: logánú a gcuid seirbhísí agus suíomhanna gréasáin; uirlisí profála, léitheoirí gréasáin agus áiseanna don Ghaeilge a eascraíonn as taighde an Phlean a bheith ar fáil ina gcuid earraí; a gcuid bolscaireachta sna meáin shóisialta a dhéanamh trí Ghaeilge (chomh maith le Béarla); urraíocht a dhéanamh ar imeachtaí ina scaiptear eolas ar theicneolaíocht na Gaeilge i measc an phobail agus i measc daltaí scoile.

16.11 An Plean Digiteach mar eiseamláir do theangacha atá i mbaol

Cé gur mionteanga atá i mbaol í an Ghaeilge, is mór an buntáiste aici go bhfuil aitheantas aici mar an chéad teanga náisiúnta, agus tacaíocht aici ón Stát – mar a fheictear i gcur chun cinn an Phlean Dhigitigh. Sa cháipéis seo cuirtear an bhéim ar fhorbairt teicneolaíochta a dhéantar le comhpháirtíocht an phobail, forbairt atá fréamhaithe i dtuiscint ar struchtúr

na teanga, atá oiriúnach i gcomhthéacs na tíre seo, ina ndírítear ar na dúshláin don Ghaeilge sa lá atá inniu ann – é sin uile chun go mbeidh na teicneolaíochtaí, na háiseanna agus na hacmhainní ag tacú le pobal labhartha na Gaeilge agus ag caomhnú na teanga mar ghnáth-theanga chumarsáide.

Cé go bhfuil an bhéim sa Phlean ar chomhthéacs na tíre seo, ní chiallaíonn sé sin nach bhfuil sé ábhartha do mhionteangacha eile atá i mbaol: meastar gur den ríthábhacht é forbairt a dhéanamh ar theicneolaíochtaí do na mionteangacha ar bhealaí a thugann aird ar struchtúr na teanga agus ar na dúshláin áitiúla, ionas go bhfuil fíoréifeacht acu ar chaomhnú teanga. Cé go bhfuil struchtúr dá cuid féin ag gach teanga, agus comhthéacs sóisialta faoi leith ag gach mionteanga, mar sin féin tá mórán dúshlán i gcoiteann acu. Níl neart go cur le chéile: nuair a aimsítear do theanga amháin bealaí le dul i ngleic le dúshlán áirithe, is eiseamláir do theangacha eile í. San fhorbairt go dtí seo ar theicneolaíocht na Gaeilge, fuarthas an-tacaíocht ó thaighdeoirí i dtíortha eile. De réir mar a chuirfear an Plean Digiteach i bhfeidhm, táimid dóchasach go mbeidh eispéireas na hÉireann ina eiseamláir do mhionteangacha eile atá i mbaol, agus go roinnfear ár dtaithí, ár saineolas, ár modhanna oibre agus nuair is féidir, ár n-áiseanna, le grúpaí eile atá ag saothrú lena dteanga féin a chaomhnú.

Tagairtí:

[1] An Straitéis 20 Bliain don Ghaeilge 2010 – 2030, An Roinn Cultúir, Oidhreachta agus Gaeltachta

[2] Ní Chiaráin, N., & Ní Chasaide, A. (2015). Evaluating synthetic speech in an Irish CALL application: influences of predisposition and of the holistic environment, in Proceedings of SLaTE 2015: 6th Workshop on Speech and Language Technologies for Education, Leipzig, Germany.

[3] Official Languages Act, available at: http://www.irishstatutebook.ie/eli/2003/act/32/enacted/en/html

[4] The Action Plan for the Irish Language, 2018-2022, An Roinn Cultúir, Oidhreachta agus Gaeltachta

[5] Disability Act, 2005, available at: http://www.irishstatutebook.ie/eli/2005/act/14/enacted/en/html

[6] The Schools' Collection, available at: www.duchas.ie

[7] www. Meitheal Logainm.ie

[8] Gmail localisation for Irish: www.thejournal.ie › gmail-as-gaeilge-1828979-Dec2014

[9] WhatsApp localisation for Irish, see Ríomhacadamh – Teichttp://teicnangael.com › riomhacadamh

[10] The MíleGlór crowdsourcing platform: https://www.abair.tcd.ie/studio/ga/recorder/

[11] McGuirk, R. (2015). Exploration of the Use of Irish Language Synthesis with a Screen Reader in the Teaching of Irish to Pupils with Vision Impairment, Unpublished MPhil thesis, School of Linguistic, Speech and Communication Sciences, Trinity College, Dublin, Ireland.

[12] European Language Resource Coordination (ELRC), see www.lr-coordination.eu

[13] CLARIN - European Research Infrastructure for Language Resources and Technology, see https://www.clarin.eu

[14] Meta-Share network of repositories for sharing language data: http://www.meta-share.org

[15] SIGUL: Special Interest Group: Under-resourced Languages, a joint special-interest group of the European Language Resources Association (ELRA) and of the International Speech Communication Association (ISCA), www.elra.info/en/sig/sigul/

AGUISÍN A

San aguisín seo, tugann muid cur síos níos cuimsithí ar na corpais éagsúla atá forbartha go dtí seo don Ghaeilge (aibhsithe i dTábla 3.1 agus i dTábla 3.2) agus ar na riachtanais láithreacha a aithníodh i dtéarmaí forbairt amach anseo.

1. Corpais Scríofa

1.1 Nua-Chorpas na hÉireann – Gaeilge

Nua-Chorpas na hÉireann – Gaeilge – acmhainn teanga é seo atá comhdhéanta de 30 milliún focal de théacsanna scríofa a chuimsíonn stíleanna agus foinsí éagsúla, comhdhéanta de leabhair, nuachtáin, irisí, téacsleabhair, reachtaíocht agus suíomhanna idirlín arna bhfíorú de láimh [1]. Is leagan leathnaithe é an corpas seo de Chorpas na Gaeilge níos luaithe a forbraíodh in ITÉ. Rinneadh anótáil uathoibríoch ar an gcorpas le clibeanna roinne cainte (m.sh. ainmfhocal, briathar, dobhriathar srl. agus gnéithe morfa-chomhréireacha eile amhail tuiseal, uimhir, inscne, aimsir, pearsa srl.) ag úsáid clibeálaí roinne cainte riailbhunaithe [2] (cruinneas os cionn 96% ar an meán).Tá meiteashonraí mionsonraithe sa chorpas.

Forbairtí Riachtanacha Amach Anseo

- Chun úsáid reatha na teanga a léiriú agus chun an iliomad míreanna léacsacha a tháinig isteach sa teanga ó chruthaíodh na sonraí seo (1995-2003) a ghabháil, tá sé riachtanach sonraí nua a chur leis an gcorpas.
- Tá corpas 30 milliún focal measartha beag de réir caighdeáin idirnáisiúnta. Tuairiscíonn Foras na Gaeilge gur gá é a mhéadú go suntasach chun freastal ar riachtanais an fhoclóra Gaeilge-Béarla reatha atá bunaithe ar chorpas.
- Tá gné labhartha an chorpais seo beag go leor (240k focal) [3] agus ní mór é a mhéadú go suntasach chun stór focal canúnach agus an iliomad gnéithe de dioscúrsa nach bhfuil le fáil i gcorpais scríofa a ghabháil.

1.2 Corpais Focal agus Nathanna Cainte ar Chaighdeán Óir

Cruthaíodh corpas ar chaighdeán óir clibeáilte de réir na roinne cainte (3,000 abairt a roghnaíodh go randamach as corpas an NCI) chun cruinneas an chlibeálaí roinne cainte riailbhunaithe críochstaide don Ghaeilge a mheas (féach Caibidil 5). Úsáideadh é freisin mar shonraí oiliúna in algartaim meaisínfhoghlama chun gnéithe moirfeolaíocha agus aicmí leamúcháin (ceannfhocal) a fhoghlaim [4]. Úsáideadh é freisin chun clibeálaí spleáchas gramadaí riailbhunaithe agus parsálaí smután a fhorbairt [5] agus cuireann sé an bonn ar fáil chun Bunachar Crainn Spleáchas Gaeilge [6, 7], a bhfuil cur síos orthu sa chéad chuid eile, a fhorbairt.

Mar sin féin, i gcomparáid leis an gcorpas is mó úsáid ar a bhfuil clibeáil ar chaighdeán óir déanta de réir na roinne cainte don Bhéarla, m.sh. Penn Treebank [8], a bhfuil breis agus 4.8 milliún focal ann le clibeanna roinne cainte ceartaithe de láimh, is lú i bhfad an corpas reatha ar a bhfuil clibeáil ar chaighdeán óir déanta de réir na roinne cainte don Ghaeilge (75K focal), rud a léiríonn an gá láithreach le leathnú.

Forbairtí Riachtanacha Amach Anseo

Cé gur tagairtí saibhre iad corpais reatha ar chaighdeán óir chun anailís a dhéanamh ar fheiniméin theangeolaíocha na Gaeilge, ní oireann a méid beag do mhodhanna staitistiúla PTN mar shamhaltú teanga nó meaisínfhoghlaim – modhanna a ligeann do ríomhaire patrúin teanga a fhoghlaim trí bhreathnú agus ríomh dóchúlachta a dhéanamh ar ghnéithe teangeolaíocha. Maidir leis na cineálacha cur chuige seo chun oibre, tá gá le corpais i bhfad níos mó. Tá sé léirithe ag méid an bhunachair crainn Penn, mar shampla, go bhfuil an-rath ar an gcur chuige seo. Mar sin, ní mór na céimeanna seo a leanas a ghlacadh maidir le corpais Ghaeilge ar a bhfuil clibeáil déanta de réir na roinne cainte:

- Méadú ar mhéid an chorpais Gaeilge ar a bhfuil clibeáil ar chaighdeán óir déanta de réir na roinne cainte.
- I gcásanna inar féidir, ba chóir go n-éascófaí tascanna lipéadaithe le modhanna brógstrapála leath-uathoibríocha, agus fíorú láimhe ina dhiaidh sin.
- Tá gá freisin le forbairt ar chorpais ar a bhfuil clibeáil déanta de réir na roinne cainte de théacs ó réimsí breise (m.sh. an Vicipéid, sonraí gréasáin, sonraí meán sóisialta, ailt eolaíocha, corpais foghlaimeoirí, srl).

1.3 Bunachar Crainn Spleáchas Gaeilge

Ó thaobh na Próiseála Teanga Nádúrtha de, feidhmíonn bunachar crainn (féach Cuid 3.4.1) mar shonraí ónar féidir le ríomhairí patrúin teanga a fhoghlaim chun parsálaí staitistiúil a chruthú agus ina dhiaidh sin próiseáil a dhéanamh ar struchtúr gramadaí abairtí Gaeilge nach bhfacthas riamh cheana. Is acmhainn luachmhar é parsálaí (féach Caibidil 5) le haghaidh feidhmeanna amhail córais meaisínaistriúcháin staitistiúil, córais freagartha ceisteanna, uirlisí achoimrithe agus seiceálaithe gramadaí. Mar sin féin, tá na bunachair chrainn Ghaeilge reatha measartha beag.

Faoi láthair tá 1020 abairt (crann párséla) i mBunachar Crainn Spleáchas na Gaeilge (BCSG) [6]. Tá treoirlínte anótála ag gabháil leis an mbunachar crainn seo ina ndéantar cur síos soiléir ar anailís spleáchas ar an nGaeilge agus treoracha le leanúint ag anótálaithe amach anseo.

Tá an méid reatha (gan ach 23k comhartha ann) i bhfad níos lú ná formhór na mbunachar crainn caighdeánach eile (m.sh. Araibis – 242k comhartha; Seicis – 1,503k comhartha; Spáinnis – 423k comhartha; Sean-Ghréigis – 244k comhartha).

Is leagan den BCSG é Bunachar Crainn Spleáchas Uilíoch na Gaeilge (BCSUG) atá tiontaithe go scéim anótála "uilíoch" mar chuid den Tionscadal um Spleáchais Uilíocha [7]. San eisiúint is déanaí, c2.10 , bhí 4,910 bunachar crainn anótáilte ó thaobh comhréire (40k comhartha). Sa tionscadal seo déantar forbairt ar bhunachair crainn teanga éagsúla (220+ faoi láthair), a úsáidtear mar bhunús do pharsáil agus anailís thras-teangach. Cuireann an tionscadal seo ar chumas teangacha ar bheagán acmhainní sonraí bunachair crainn agus faisnéis theangeolaíoch a ghiaráil ó theangacha a bhfuil níos mó acmhainní acu, mar mhodh chun easpa sonraí a shárú. Agus an BCSUG san áireamh sa tionscadal UD, tá ardú tagtha ar phróifíl na Gaeilge i staidéir próiseála teanga nádúrtha agus thug sé deis do thaighdeoirí eile freisin ár dtacair sonraí a úsáid i staidéir thrasteangacha parsála nó thíopeolaíochta. Tá na buntáistí a bhaineann le scaoileadh na sonraí Gaeilge le sonrú cheana féin, toisc go bhfuil an-suim sa teanga i roinnt staidéar mar gheall ar a gnéithe uathúla teanga, rud a chuireann dúshláin nua os comhair an phobail seo [9]. I saothair eile tugadh tuairisc ar chuimsiú na Gaeilge in anailís thrasteangach ar theangacha atá ar bheagán acmhainní [10] chomh maith le tascanna comhroinnte párséla [11].

Chun acmhainn dá leithéid a dhearadh agus a thógáil, teastaíonn corpas ar a bhfuil clibeáil ar chaighdeán óir déanta de réir na roinne cainte, anailís fhairsing theangeolaíoch agus anótálaithe atá oilte sa teanga agus a thuigeann na modhanna ríomhaireachta don pharsáil. Tógann an tasc am – ag brath ar chastacht abairte, d'fhéadfadh idir 5 nóiméad agus uair an chloig a bheith de dhíth chun abairt a anótáil.

Forbairtí Riachtanacha Amach Anseo
Tá cruinneas reatha 71.4% (3) ag parsálaí spleáchas na Gaeilge agus tá cruinneas 73.1% [12] (84% i gcás gaBERT [35]) ag parsálaí spleáchas uilíoch na Gaeilge (i gcomparáid le parsálaithe spleáchas uilíoch den chineál is nua-aimseartha don Fhraincis 91%,93% don tSeicis agus 89% don Bhéarla). Tá comhchoibhneas díreach idir feidhmíocht parsálaí agus méid bunachair crainn. Tá gá le bunachar crainn i bhfad níos mó, mar sin, chun iarrachtaí parsála staitistiúla amach anseo a éascú, chun caighdeán feidhmeanna iartheachtacha eile a chinntiú (mar shampla meaisínaistriúchán, aisghabháil faisnéise, seiceáil gramadaí, foghlaim ríomhchuidithe teangacha, achoimriú téacs).

- Ba chóir go mbeadh an leathnú ar an mBunachar Crainn Spleáchas Gaeilge (BCSG) ag tarlú i gcomhthráth leis an leathnú ar an gcorpas ar a bhfuil clibeáil ar chaighdeán óir déanta de réir na roinne cainte.
- Le forbairt an Bhunachair Crainn Spleáchas Uilíoch (BCSUG) ba chóir go gcomhlíonfaí eisiúintí bliantúla de réir sceideal an tionscadail UD ag cinntiú go leantar ag úsáid sonraí Gaeilge i staidéir ar fud an domhain.
- Ba chóir úsáid a bhaint as modhanna brógstrapála nuair is féidir chun forbairt na n-acmhainní seo a bhrostú.
- Ba chóir cur chuige sluafhoinsithe (féach Caibidil 16) a fhiosrú i bhforbairt bunachair crainn (m.sh. modhanna cluichithe a úsáid chun an foghlaim teanga, cluichíocht agus tógáil acmhainní a thabhairt le chéile chun an bunachar crainn a anótáil, mar a rinneadh i gcás anótáil na Fraincise).
- Tá easnamh in BCSG ó thaobh anótáil chuimsitheach ar nathanna ilfhoclacha, ar struchtúir chomhréire iad a bhfuil ról ríthábhachtach acu i bpróiseáil chruinn teanga [13, 14]. De thoradh na gcéimeanna tosaigh den fhorbairt ar fhoclóir ilfhoclach don Ghaeilge [15] tá bonn taca curtha faoin taighde sa réimse seo [34]. Tá tuilleadh oibre ag teastáil sa réimse seo i gcomhréir leis na gníomhaíochtaí faoi Ghníomh UniDrive EU Cost.

1.4 Corpas Gréasáin Gaeilge Crúbadán

Is éard atá i gCorpas Gaeilge Crúbadán, a forbraíodh mar chuid den Tionscadal Crúbadán [16], ná corpas cuimsitheach de théacsanna Gaeilge a ransaíodh ón ngréasán lena n-áirítear ailt nuachta, blagphostálacha, postálacha ar na meáin shóisialta, cáipéisí rialtais agus dlí, litríocht, srl. Tar éis glantacháin, abairtí dúblacha agus téacs scríofa i dteangacha eile a bhaint, tá breis agus 100 milliún focal Gaeilge sa chorpas. Tá an corpas seo tábhachtach i bpáirt mar gheall ar a mhéid, rud a éascaíonn forbairt teicneolaíochtaí teanga a bhraitheann ar shamhaltú staitistiúil teanga.

Tiomáineann corpas an Chrúbadáin an fhorbairt ar roinnt teicneolaíochtaí tábhachtacha teanga, ina measc, litreoir GaelSpell , an seiceálaí gramadaí An Gramadóir , na hinnill mheaisínaistriúcháin InterGaelic ó Ghaeilge Mhanann agus ó Ghaeilge na hAlban go Gaeilge [17] agus an Caighdeánaitheoir Gaeilge [18]. Tá fo-thacair den chorpas gréasáin in úsáid freisin ag na foclóirithe a bhí ag obair ar thionscadal an Fhoclóra Nua Béarla-Gaeilge agus ag téarmeolaithe an Choiste Téarmaíochta mar mhodh chun focail nuachumtha Ghaeilge a aithint.

Forbairtí Riachtanacha Amach Anseo
Tá sé tábhachtach go leantar den chorpas ransaithe seo a nuashonrú go laethúil de réir mar a thagann téacsanna nua chun cinn ar an ngréasán, agus mar sin go dtugann sé léargas cruinn ar an teanga mar a úsáideann cainteoirí an lae inniu í.

1.5 An Vicipéid: Ciclipéid Ghaeilge ar Líne

Is ciclipéid ilteangach ar líne í an Vicipéid a forbraíodh

trí Fhondúireacht Wikimedia. Gné thábhachtach de sholáthar acmhainní téacs teanga is ea fás na Vicipéide Gaeilge. Ní hamháin go gcuireann Vicipéid na Gaeilge sonraí méadaithe téacs ar fáil do thaighde agus d'fhorbairt PTN, ach freisin, trí na meiteashonraí breise éagsúla a chuirtear ar fáil tríd an struchtúr vicí, (eochairfhocail, hipearnaisc, srl.), is bonn luachmhar é a leithéid de chorpas freisin do thascanna próiseála teanga amhail aithint eintiteas ainmnithe (Caibidil 5), anailís ar sheintimint (Caibidil 5), giniúint teanga nádúrtha (Caibidil 6), agus mar sin de. Baineann buntáiste oideachasúil freisin le foghlaimeoirí teanga trí thagairtí téacs a chruthú ar éagsúlacht topaicí agus seánraí Gaeilge. Ar deireadh, is féidir le cainteoirí dátheangacha leas a bhaint as rochtain a fháil ar an stór eolais seo a úsáidtear go huilíoch trí Ghaeilge, seachas rochtain a fháil ar an ábhar trí Bhéarla amháin.

Tá an Vicipéid Ghaeilge sa 93ú háit ar an liosta faoi láthair de gach vicipéid teanga, agus 57,000 alt ann. Chun é seo a chur i gcomhthéacs, tá an Bhreatnais sa 41ú háit agus tá an Afracáinis (68ú), an Astúiris (64ú) agus an Bhriotáinis (82ú) i measc na dteangacha a bhfuil rangú níos airde acu ná mar atá ag an nGaeilge.

Forbairtí Riachtanacha Amach Anseo

- Tá go leor oibre ag teastáil chun líon na n-eagarthóirí Gaeilge Vicipéide a mhéadú agus líon na n-alt sa Vicipéid a mhéadú.
- Teastaíonn comhar agus comhoibriú idir geallsealbhóirí éagsúla amhail Pobal Wikimedia Éireann, Fondúireacht Wikimedia, eagraíochtaí oideachais (ollscoileanna agus scoileanna), pobal na próiseála teanga nádúrtha agus saineolaithe Gaeilge chun cur chuige straitéiseach a chomhordú chun an Vicipéid a fhorbairt chun acmhainn ardchaighdeán atá ar fáil go poiblí a chinntiú. I measc na múnlaí rathúla de leathnú na Vicipéidí ó phobail teanga eile amhail an Bhreatnais , an Bhriotáinis, agus an Bhascais, tá oiliúint shainiúil ar dhaltaí meánscoile, agus meaisínaistriúchán a dhéanamh ar ábhar ó Vicipéidí teangacha eile (agus iareagarthóireacht le déanamh ag eagarthóirí air ina dhiaidh sin).

1.6 Corpas Gaeilge Comhaimseartha Gaois

Is éard atá i gCorpas Gaeilge Comhaimseartha Gaois (25+ milliún focal) ná bailiúchán de théacsanna Gaeilge Fiontar agus Scoil na Gaeilge (OCBÁC) i bhformáid dhigiteach atá inrochtana le haghaidh fiosruithe nó taighde ar Gaois.ie. Is éard atá ann ná téacsanna curtha in eagar a foilsíodh ó thús an 21ú céad ar aghaidh. Áirítear sa chorpas faoi láthair ailt sna meáin, píosaí litríochta agus scríbhneoireacht acadúil. Cuireadh sonraí ceanntásca leis an gcorpas ar fad de láimh maidir leis an bhfoilsitheoir agus seánra an téacs, agus tá thart ar 40% de seo marcáilte suas mar XML de réir chaighdeáin an TEI (Tionscnamh um Ionchódú Téacs) . Tá tús curtha le hobair ar anótáil theangeolaíoch an chorpais le haghaidh clibeanna de réir na roinne cainte, leimiúchán agus faisnéis mhoirfeolaíoch, ach is ag céim luath atá sé. Bailítear téacsanna breise ó am go ham agus cuirtear leis an gcorpas iad. Tá foinsí nua sonraí á lorg go gníomhach ag Grúpa Taighde Gaois.

Forbairtí Riachtanacha Amach Anseo

- Tá timpeall 25 milliún focal sa chorpas faoi láthair ach ní mór é a leathnú ar bhealach cothrom chun a chinntiú go gcoimeádtar cothrom le dáta é le húsáid reatha na teanga.
- Teastaíonn anótáil theangeolaíoch ar nós clibeanna de réir na roinne cainte, leimiúchán agus faisnéis mhoirfeolaíoch freisin chun úsáid leathan a chinntiú i dteicneolaíocht teanga agus i dtaighde teangeolaíoch.

I gcás ina gcuireann srianta cóipchirt cosc ar shonraí corpais a scaoileadh go poiblí, ba chóir gach iarracht a dhéanamh fothairgí na n-acmhainní sin a sholáthar, mar shampla leabuithe focal réamhoilte nó samhlacha teanga.

1.7 Corpas Comhthreomhar Gaois

Corpas comhthreomhar de mhíreanna téacs ailínithe ó reachtaíocht na hÉireann agus an Aontais Eorpaigh. 26 milliún focal Gaeilge agus 24.5 milliún focal Béarla.

1.8 Corpais de Théacsanna Meán Sóisialta Gaeilge

Tá 1,500 tvuít randamach i gcorpas tvuíteanna Gaeilge. Baineadh úsáid as an gcorpas seo chun Bunachar Crainn Spleáchas TwitterIrish a thógáil [19], agus leas á bhaint astu mar chorpais traenála le haghaidh clibeálaithe roinne cainte tvuít-sonracha agus le haghaidh parsálaithe spleáchais. Is féidir uirlisí den sórt sin a chomhtháthú i bhfeidhmeanna iartheachtacha agus dá thoradh sin anailís a dhéanamh ar shonraí Gaeilge Twitter [4, 19]. Ina theannta sin, leathnaíodh an dá chorpas seo le hanótálacha maidir le códmhalartú Béarla [19, 20]. Tá corpas beag de 1,000 tvuít chomhthreomhar Béarla-Gaeilge a bhaineann le Corn an Domhain sa Bhrasaíl 2014 [21] bailithe freisin ag taighdeoirí in Ollscoil Chathair Bhaile Átha Cliath. Tá na tvuíteanna sin ina dtús le corpas is féidir a úsáid chun córais a thógáil chun anailís a dhéanamh ar ábhar Gaeilge Twitter a bhaineann leis an spórt.

Forbairtí Riachtanacha Amach Anseo

Tá an obair atá ar bun sa réimse seo cheana féin ina bhunús do léann na sochtheangeolaíochta ar an leagan nua den Ghaeilge atá ag teacht chun cinn ar líne, agus cuireann sé léargais ar fáil do ghlúin nua cainteoirí Gaeilge, agus ar na bealaí éagsúla inar féidir le daoine foghlaim agus dul i ngleic leis an teanga. Is réimse taighde é seo a mbeidh staidéar i bhfad níos fairsinge ag teastáil lena aghaidh.

- Beidh gá le leathnú agus nuashonrú suntasach ar chorpais na meán sóisialta chun úsáid reatha na teanga ar líne a léiriú.
- Tá gá le tuilleadh taighde ar úsáid an chódmhalartaithe sna meáin shóisialta idir Béarla agus Gaeilge ar líne, agus mar sin beidh gá le corpas níos mó.
- I gcás staidéir a dhéanfar amach anseo ar úsáid haischlibeanna nó ar chatagóiriú tvuíteanna Gaeilge, beidh gá le cineálacha sonracha tvuíteanna a bhailiú seachas tvuíteanna ó fhearainn ghinearálta.
- Tá gá le tuilleadh taighde chun cruinneas na clibeála roinne cainte Twitter a fheabhsú.
- Tá gá le tuilleadh taighde ar pharsáil (asbhaint struchtúr abairtí) Tvuíteanna.
- Ina theannta sin, tá gá le hiniúchadh a dhéanamh ar úsáid na Gaeilge ar Facebook agus ar líonraí sóisialta eile dá leithéid (is é sin, ní hamháin ar Twitter).
- I gcás uirlisí chun anailís a dhéanamh ar sheintimint an ábhair ar líne, teastaíonn iad a chur in oiriúint don Ghaeilge chun a chinntiú go gcuirtear san áireamh go cothrom na tuairimí a chuireann pobal na Gaeilge in iúl ar líne; tá réamhstaidéir déanta ina leith [22].

1.9 Corpais de Nathanna Cainte Ilfhoclacha Clibeáilte

I nathanna ilfhocalacha (NIFanna) bíonn sraith dhá fhocal nó níos mó a chuireann brí amháin in iúl (m.sh. buail le [lit. hit with] 'meet'; déan iarracht 'make an attempt'; cuir fearg (ar) [lit. put anger (on)] 'make angry'; bain amach [lit. extract out] 'earn'). Cruthaíonn na nathanna seo dúshláin do phróiseáil teanga nádúrtha (Caibidil 5) mar gur leaganacha cainte iad agus toisc gur deacair ciall a bhaint astu de réir na bhfocal aonair amháin.

Tá tús curtha le taighde i réimse na próiseála uathoibríoch ar NIFanna na Gaeilge le scéim chun catagóiriú NIFanna a fhorbairt mar aon le cruthú corpas NIF-chlibeáilte ar a mbunófar sonraí tástála agus traenála le haghaidh uirlisí atá feasach ar NIFanna [38]: (i) Is éard atá i gcorpas NIFanna Gaeilge PARSEME ná 1,700 abairt a bhí i mbunachair crainn UD na Gaeilge ar dtús [7], agus iad anótáilte le 662 vNIF ó sheacht gcatagóir: nathanna briathartha, briathra a bhfuil feidhm réamhfhoclach nó iarfhoclach acu, briathra a bhfuil feidhm aisfhillteach acu, struchtúir briathair éadroim, agus struchtúir míre briathartha. (ii) Chun iniúchadh a dhéanamh ar an gcaoi a n-oibríonn NIFanna i meaisínaistriúchán Gaeilge-Béarla agus Béarla-Gaeilge agus ar an toradh a bhíonn orthu, cruthaíodh corpas comhthreomhar NIFanna (as 835,867 abairt ailínithe go dátheangach) agus rinneadh é a chlibeáil go huathoibríoch le NIFanna don Bhéarla agus don Ghaeilge araon. [iii] Bunaíodh corpas NIFanna ar chaighdeán óir don Ghaeilge agus don Bhéarla trí 25 abairt ó gach ceann de na ceithre réimse a léirítear sna sonraí a chomhcheangal: an fearann teicniúil, an fearann dlíthiúil, an fearann gréasáin agus an fearann ginearálta.

1.10 EduGA: Corpas d'Ábhair Oideachais i nGaeilge

Mar gheall go mbraitheann cuid mhaith de dhaonra na hÉireann ar an gcóras oideachais chun an Ghaeilge a fhoghlaim [23] tá an cineál teanga a úsáidtear agus a mhúintear sa réimse seo an-tábhachtach do thodhchaí na Gaeilge. Tá corpas 7.5 milliún focal de théacsleabhair agus d'ábhair oideachais ó gach leibhéal oideachais curtha le chéile agus anótáilte mar chuid de thaighde PhD i gColáiste na Tríonóide, Baile Átha Cliath. Dírítear sa taighde seo ar bhearta castachta teanga le haghaidh réimse an oideachais, chun castacht na teanga a úsáidtear in ábhair i measc na leibhéal éagsúil oideachais a chur i gcomparáid lena chéile (m.sh. bunleibhéal agus iar-bhunleibhéal), agus freisin chun anailís a dhéanamh ar úsáid teanga in ábhair a mhúintear as Gaeilge [24].

Forbairtí Riachtanacha Amach Anseo

- Teastaíonn nuashonrú leanúnach ar an gcorpas seo chun na riachtanais oideachais reatha agus an tírdhreach éabhlóideach san úsáid teanga i réimsí oideachais a léiriú.
- Chun gur féidir le cruthaitheoirí ábhar (lena n-áirítear foilsitheoirí tráchtála agus múinteoirí araon) leas a bhaint as an taighde seo ní mór API, nó aip ghréasáin, a chur ar fáil. D'fhéadfadh cruthaitheoirí ábhar ceacht a chruthaigh siad a ionchur, a phróiseáiltear ansin ar roinnt bealaí, agus cuirtear aiseolas oibiachtúil ar fáil.
- Tá gá le hanailísí comparáideacha idir teanga a ghineann foghlaimeoirí agus a bhfuil sna hábhair óna bhfoghlaimíonn siad. Mar shampla, d'fhéadfaí anailís a dhéanamh ar chomhréir, ar ghramadach agus ar stór focal na dtéacsleabhar seachas ar aistí na bhfoghlaimeoirí, chun ceachtanna fócasaithe a chur ar fáil a bheadh in oiriúint do riachtanais grúpa áirithe foghlaimeoirí.
- Forbairt bhreise ar an méadracht chastachta agus inléiteachta a forbraíodh sa taighde seo ionas go mbeidh an Ghaeilge ar comhchéim le teangacha Eorpacha eile.

1.11 Corpas Foghlaimeoirí Gaeilge Scríofa agus Labhartha TEG

Tá Corpas Foghlaimeoirí Gaeilge á dhearadh, á bhailiú agus á anótáil faoi láthair mar chuid de thaighde PhD san Ionad do Léann na Teanga agus na Cumarsáide, Coláiste na Tríonóide, Baile Átha Cliath, i gcomhar le Lárionad na Gaeilge – Taighde, Teagasc agus Tástáil in Ollscoil Mhá Nuad. Is éard atá sa chorpas faoi láthair ná 200,000 focal de shonraí labhartha agus scríofa foghlaimeoirí, ó scrúduithe Theastas Eorpach na Gaeilge, ag leibhéil meánleibhéil agus ardleibhéil. Déantar an corpas a thras-scríobh agus a anótáil de láimh i dtaca le hearráidí agus ceartúcháin teanga. Ina dhiaidh sin déantar anótáil uathoibríoch le clibeanna de réir na roinne cainte [25]. Táthar ag scrúdú freisin brath earráidí leath-uathoibríoch. Déanfar anailís ar an

gcorpas seo chun tuairisceoirí teanga-shonracha níos soiléire a sholáthar ar leibhéil oilteachta CTET ó B1 go C1. Tacóidh an anailís sin freisin le forbairt ábhar teagaisc, acmhainní foghlama agus uirlisí measúnaithe eimpíreach-bhunaithe atá oiriúnaithe do riachtanais na bhfoghlaimeoirí ag gach leibhéal oilteachta.

Forbairtí Riachtanacha Amach Anseo

- Soláthróidh an corpas seo bonn modheolaíochta d'fhorbairt agus d'anailís ar shonraí corpais foghlaimeoirí sa Ghaeilge, ach teastaíonn leathnú suntasach ó thaobh méide, aois na bhfoghlaimeoirí agus leibhéal oilteachta, agus seánra an téacs d'fhonn staidéar cuimsitheach ar theanga foghlaimeoirí sa Ghaeilge a éascú.
- Tá sé riachtanach go háirithe go gcuirfí tuilleadh sonraí foghlaimeoirí labhartha, go háirithe idirghníomhú nádúraíoch (m.sh. teanga an tseomra ranga agus an mhúinteora) san áireamh i gCorpas na bhFoghlaimeoirí.
- Tá gá le tuilleadh oibre ar uathoibriú earráidí a bhrath agus ar anótáil ionas gur féidir an corpas foghlaimeoirí seo a mhéadú, chomh maith le forbairt uirlisí ceartaithe earráide spriocdhírithe do mhúinteoirí agus d'fhoghlaimeoirí na Gaeilge.

1.12 Corpas náisiúnta meaisínaistriúcháin agus Corpas Comhthreomhar CEAT

I gcomparáid le teangacha oifigiúla eile an Aontais Eorpaigh, tá teorainn leis na tacair sonraí comhthreomhara atá ar fáil go poiblí don Ghaeilge. In 2015, rinneadh na chéad chéimeanna i dtreo sonraí den chineál sin a bhailiú ar bhonn náisiúnta san fhorbairt a rinne Ollscoil Chathair Bhaile Átha Cliath ar an gcóras Tapadóir MAS (Meaisínaistriúchán Staitistiúil) don Roinn Turasóireachta, Cultúir, Ealaíon, Gaeltachta, Spóirt agus Meán [26]. Áirítear leis seo corpas gréasán-ransaithe [16], doiciméid dlí Éireannacha, sonraí gréasáin nua-ransaithe agus comhaid cuimhne aistriúcháin ón Roinn. Ó shin i leith tá an córas nuashonraithe le sonraí ó fhoinsí eile ar nós Údarás na Gaeltachta, Oifig an Choimisinéara Teanga, comhairlí cathrach, ollscoileanna, agus roinnt ranna rialtais. Bhí sonraí aonteangacha ó Fhoras na Gaeilge luachmhar freisin do shamhaltú teanga san fhorbairt córais seo. Cabhróidh an bailiú sonraí seo freisin le forbairt aon chórais aistriúcháin le húsáid laistigh den tSeirbhís Comhroinnte Aistriúcháin náisiúnta atá beartaithe (féach Caibidil 11).

Tagann tionscadal taighde agus forbartha an Tapadóra le gníomhaíochtaí an tionscadail um Chomhordú Eorpach Acmhainní Teanga (CEAT). Is tionscnamh de chuid an Choimisiúin Eorpaigh é seo chun sonraí aistriúcháin a bhailiú ó na teangacha Eorpacha uile d'fhonn córas eTranslation an Aontais a fheabhsú. Mar bhall de CEAT, tá dhá Phointe Ancaire Náisiúnta ag Éirinn – ionadaí riaracháin phoiblí sa Roinn TCEGSM agus ionadaí teicniúil in Ionad ADAPT, Ollscoil Chathair Bhaile Átha Cliath. Mar gheall ar a rannphairtíocht siúd, spreagadh bailiú náisiúnta sonraí aistriúcháin ón riarachán poiblí chun uirlisí aistriúcháin a fheabhsú ag an leibhéal Eorpach agus náisiúnta. Dá thoradh sin, cuireadh tuarascáil le chéile maidir le bainistiú sonraí aistriúcháin agus úsáid na teicneolaíochta aistriúcháin i riarachán poiblí na hÉireann.

In 2019, le cistiú ón Aontas Eorpach, forbraíodh an Stáisiún Athsheachadta Náisiúnta (tairseach ar líne) chun gurbh fhéidir tabhairt faoin mbailiú sonraí seo. Rinneadh an tairseach a athsheoladh in 2022 faoin ainm eStór. Tugtar rochtain ar an tairseach sonraí seo do gach duine a oibríonn le sonraí Gaeilge sa riarachán poiblí. Leis an suíomh seo, is féidir comhaid cuimhne aistriúcháin a ghiniúint ó bhaisceanna de dhoiciméid aistrithe chomhthreomhara tar éis na doiciméid sin a uaslódáil. Ina theannta sin, trí chomhroinnt ghrúpa, tá úsáideoirí in ann rochtain a fháil ar shonraí teanga (m.sh. comhaid cuimhne aistriúcháin) ó úsáideoirí eile a chabhróidh lena n-iarrachtaí aistriúcháin féin, agus in go leor cásanna laghdófar costais aistriúcháin amach anseo. Tá formhór na sonraí aistriúcháin a uaslódáil daoine atá i riarachán poiblí ar fáil saor in aisce le híoslódáil faoi Threoir Sonraí Oscailte an AE, le tacaíocht ón Aonad Náisiúnta Sonraí Oscailte. Tá an chuid is mó de na sonraí aistriúcháin arna n-uaslódáil ag daoine sa riarachán poiblí ar an tairseach ar fáil le híoslódáil saor in aisce faoi Threoir an Aontais Eorpaigh maidir le Sonraí Oscailte, le tacaíocht ón Aonad Sonraí Oscailte náisiúnta.

I dtionscadal ina dhiaidh sin a mhaoinigh an tAontas Eorpach agus a reáchtáladh in DCU, PRINCIPLE, díríodh ar shonraí a bhailiú i gcomhair teangacha atá, dar leis an gCoimisiún Eorpach, "ar bheagán acmhainní" ó thaobh na sonraí a theastaíonn chun innill mheaisínaistriúcháin a thógáil (Gaeilge, Cróitis, Íoslainnis agus Ioruais). De thoradh an tionscadail seo bhí cúnamh le fáil ag páirtithe leasmhara in Éirinn chun sonraí aistriúcháin atá ann cheana nó sonraí cartlainne a bhailiú, a phróiseáil, a ghlanadh agus a ailíniú ionas gur féidir iad a úsáid in obair aistriúcháin agus chun teicneolaíocht an aistriúcháin a fhorbairt. I measc na ndreamanna a ghlac páirt agus a chuidigh sa tionscadal seo bhí Rannóg an Aistriúcháin, Foras na Gaeilge, Ollscoil na Gaillimhe agus an Roinn Dlí agus Cirt, agus chlúdaigh siad réimsí an r-Cheartais agus an r-Sholáthair. Is sampla den scoth é an tionscadal sin den chaoi ar féidir, de thoradh sonraí aistriúcháin dea-bhainistithe agus dea-choimeádta a chomhroinnt, feabhas a chur ar ghníomhaíochtaí aistriúcháin trí laghdú ama agus costais in eagraíocht phoiblí.

Forbairtí Riachtanacha Amach Anseo

- Ní mór tacaíocht riaracháin leanúnach a thabhairt do eStór agus cothabháil theicniúil leanúnach a dhéanamh ar an tairseach sin lena chinntiú go leantar le bailiú agus bainistiú sonraí aistriúcháin ar leibhéal náisiúnta.
- Tá gá le comhoibriú leanúnach leis an gCoimisiún Eorpach san fheachtas faoi stiúir CEAT le haghaidh bailiú sonraí (faoi chlár Aistriúcháin Uathoibrithe na

Saoráide um Chónascadh na hEorpa (SCE)) agus faoi Spásanna Sonraí Comhchoiteanna an Choimisiúin) atá dírithe ar fheabhas a chur ar aistriúchán uathoibrithe sa riarachán poiblí ar fud na dtíortha AE uile (m.sh. Seirbhísí Aistriúcháin Comhroinnte Rialtas na hÉireann).

- Tá géarghá freisin le feabhas a chur ar an ransú sonraí gréasáin do shonraí Béarla-Gaeilge. Tá gá le comhoibriú níos dlúithe le forbróirí ransaitheoirí gréasáin (m.sh. ransaitheoir gréasáin ILSP [16, 27] chun modhanna a fheabhsú chun téacsanna Béarla-Gaeilge a chamchóipeáil ó fhearainn .ie. Mar sin féin, tá inneachar meaisínaistrithe i gcorpas ransaithe Gaeilge an tionscadail Paracrawl [28] (19) mar aon le go leor sonraí a bailíodh cheana féin trí mhodhanna CEAT.
- Tá gá freisin le modhanna a fhiosrú chun frásaí ailínithe ábhartha a bhaint as corpais inchomparáide le cur san áireamh i dtacair sonraí comhthreomhara téacs. D'fhéadfaí féachaint air seo mar iarracht chun an easpa acmhainní corpas Gaeilge a shárú agus an méid téacs atá ar fáil i ndáiríre a uasmhéadú.
- Teastaíonn comhoibriú níos dlúithe le hAonad Sonraí Oscailte na Roinne Caiteachais Phoiblí agus Athchóirithe agus leis an Oifig um Sholáthar Rialtais chun feasacht a chruthú ar luach sonraí aistriúcháin agus ar an tairbhe atá le baint ag an Stát as gníomhaíochtaí comhroinnte sonraí a fheabhsú.

1.13 Cnuasach Bhéaloideas Éireann – Bailiúchán na Scol

Mar chuid den tionscadal Meitheal Dúchas.ie chun Cnuasach Bhéaloideas Éireann a dhigitiú, tá lámhscríbhinní Bhailiúchán na Scol á dtras-scríobh ag an bpobal tríd an ngréasán. Cuimsíonn torthaí na hoibre seo anois corpas sluafhoinsithe de bhreis agus 3.8 milliún focal i nGaeilge ó na 1930idí ar féidir cuardach a dhéanamh air trí Dúchas.ie. Fuair an suíomh gréasáin 450k cuairteoir uathúla in 2019.

Forbairtí Riachtanacha Amach Anseo
Tá feabhas mór ag teastáil ar an áis chuardaigh ar shuíomh gréasáin Dúchas.ie. Léirigh réamhstaidéar le déanaí go bhféadfaí leas a bhaint as teicneolaíocht teanga chun feabhas a chur ar an gcaoi ar féidir an bunachar sonraí a chuardach [29].

1.14 Corpas Stairiúil na Gaeilge, 1600–2000

Is corpas cuimsitheach den Ghaeilge é Corpas Stairiúil na Gaeilge a théann siar 400 bliain. Tá an corpas á fhorbairt in Acadamh Ríoga na hÉireann mar bhunús le Foclóir Stairiúil na Gaeilge, tionscadal taighde a bhfuil sé mar aidhm aige foclóir stairiúil Gaeilge a chruthú don tréimhse 1600–2000.

Tá an corpas comhdhéanta de chúig mhór-fhochorpas ina gcuimsítear tréimhsí stairiúla éagsúla na Nua-Ghaeilge. Go dtí seo tá fo-chorpas do na tréimhsí 1600–1882 agus 1882–1926 tiomsaithe agus anótáilte agus ar fáil ar líne . Mar gheall gur corpas stairiúil é seo, agus chun sonraí a ghabháil do na tréimhsí ama níos luaithe, teastaíonn aithint optúil carachtar (AOC) agus seiceáil de láimh, mar aon le caighdeánú teanga (de láimh agus uathoibríoch) agus clibeáil uathoibríoch roinne cainte [30].

Forbairtí Riachtanacha Amach Anseo
Tá gá le hobair ar thréimhsí ina dhiaidh sin den chorpas le gur féidir Corpas Stairiúil na Gaeilge a chríochnú agus chun obair ar an bhfoclóir féin a éascú.

1.15 Corpas Fhilíocht Shiollach na Gaeilge, 1200–1650

Is éard atá sa chorpas seo beagnach dhá mhíle dán a scríobhadh i nGaeilge Chlasaiceach sa tréimhse ó 1200 go 1650. Faoi láthair tá cúig chéad de na dánta seo agus meiteashonraí fairsinge curtha ar fáil ar líne do scoláirí agus don phobal. Féach https://www.tcd.ie/slscs/research/areas/corpora/bardic.php le tuilleadh sonraí a fháil.

Forbairtí Riachtanacha Amach Anseo
Tá gá le tuilleadh taighde chun leimiú agus clibeáil roinne cainte do théacsanna na tréimhse seo a fheabhsú, chun an chuid eile den chorpas seo a chur ar fáil go poiblí.

2. Corpais Urlabhra agus Labhartha

Sa chuid seo, tugaimid achoimre ar roinnt corpais ciainte (caint scriptithe) agus corpais labhartha (caint spontáineach).

2.1 Corpas Ginearálta Sintéise Urlabhra ABAIR

Tá corpais á ndearadh agus á dtaifeadadh go leanúnach laistigh de ABAIR do ghuthanna sintéiseacha na gcanúintí Gaeilge éagsúla. Is é an prionsabal ginearálta ná ábhair atá oiriúnach don chanúint a úsáid chun fuaimeanna a thaifeadadh agus féachaint le clúdach maith foghraíochta a bhaint amach i ngach comhthéacs. Tá éagsúlacht i méid na gcorpas seo, rud a léiríonn riachtanais an innill cainte ar leith a úsáidtear chun guth a fhorbairt (féach Caibidil 8). Go dtí seo, díríodh ar na 3 phríomhchanúint ach tá obair ar bun ar fhochanúintí eile. Tá timpeall 25 uair an chloig de chaint thaifeadta ann: tá codanna roghnaithe de na corpais seo optamaithe le haghaidh sintéise – curtha in eagar, anótáilte lena gceartú le haghaidh stadanna sa chaint srl., deighleánú déanta orthu, iad ailínithe leis an trascríobh foghraíochta (X-SAMPA, IPA) le marcáil aicinn srl.

Forbairtí Riachtanacha Amach Anseo
Mar a thuairiscítear i gCaibidil 8, beidh gá le breis dearadh corpais, taifeadadh agus ullmhú/anótáil chun

cúram a dhéanamh den mhéid seo:

- Clúdach iomlán ar na canúintí Gaeilge, na cinn is mó i mbaol san áireamh. Áirítear leis seo ábhair a bhaineann go sonrach le canúint faoi leith a bhailiú agus anailís leanúnach ar chlúdach foghraíochta na gcorpas
- Tuilleadh cainteoirí, guthanna leanaí go háirithe
- Corpais le haghaidh códmhalartú agus sintéis dhátheangach ar deireadh
- Corpais atá saibhir ó thaobh prosóide a chur ar fáil
  - le haghaidh raon níos leithne feidhmeanna sintéise (guthanna nach bhfuil chomh haontonach céanna)
  - corpais chainte léiritheacha idirghníomhacha (lena n-áirítear mothúchán, giúmar, dearcadh)

Ba chóir go bhfeabhsódh na corpais seo caighdeán an aschuir cainte sintéisí uile ach tá siad ríthábhachtach do chórais don urlabhra chomhrá.

2.2 Dlúthchorpas sintéise ABAIR: an Corpas Beag

Is ionann an Corpas Beag agus sraith corpas atá á bhforbairt ag grúpa ABAIR i gColáiste na Tríonóide. Dlúthchorpais iad seo do na canúintí Gaeilge aonair agus é mar aidhm acu an clúdach foghraíochta is mó is féidir a sholáthar leis an méid is lú d'ábhar leideanna scriptithe. Tá a leithéid de chorpas deartha do chanúint na Mumhan (Corpas Beag na Mumhan) agus tá obair ar siúl ar dhlúthchorpas do chanúintí eile na Gaeilge.

Forbairtí Riachtanacha Amach Anseo
Is ionann na ceanglais agus iad siúd a thuairiscítear do na corpais ghinearálta sintéise a shonraítear in §2.1 thuas, ach go bhfuil gá freisin le:

- Ábhar a chruthú atá feiliúnach don chanúint agus a sholáthraíonn clúdach iomlán foghraíochta.
- Corpais a thaifeadadh, a anótáil agus a bheachtú le haghaidh optamú in innill chainte shonracha

2.3 Corpas Labhartha Comhrá

Faoi láthair tá 240k focal sa Chorpas Comhrá de chaint thras-scríofa spontáineach, agus cuimsíonn sé breis agus 200 cainteoir ó na mhórchanúintí Gaeilge uile. Ceapadh an corpas labhartha seo chun clúdach cothrom a thabhairt ar na canúintí agus ar ghnéithe sochtheangeolaíochta eile amhail aois agus inscne. Is ionann an corpas reatha agus fo-thacar de dhearadh an chorpais bunaidh .

Faoi láthair tá breis agus céad closchomhad tras-scríofa sa chorpas, gach ceann díobh ocht nóiméad ar fad ó fhoinsí éagsúla lena n-áirítear podchraoltaí Ráidió na Gaeltachta, closchomhaid chartlann RTÉ, agus roinnt samplaí ó Chorpas Gaeilge Labhartha Chonamara Hartmann a taifeadadh i Ros Muc, Co. na Gaillimhe sna 1960idí [31]. Ina theannta sin, tá roinnt taifeadtaí ann ar idirghníomhaíochtaí spontáineacha, a taifeadadh i gceithre shuíomh Gaeltachta agus i mBaile Átha Cliath. Le clibeálaí roinne cainte críochstaide riailbhunaithe, rinneadh clibeáil roinne cainte ar chorpas Comhrá agus feabhsaíodh an clibeálaí seo chun gnéithe a bhaineann go sonrach leis an teanga labhartha a láimhseáil [3]. Cuimsíonn an clibeálaí modúl Béarla chun míreanna léacsacha agus códmhalartú Béarla a chlibeáil go beacht. Tá an corpas reatha, atá maoinithe ag Foras na Gaeilge do thionscadal an Fhoclóra Nua Béarla-Gaeilge https://www.focloir.ie/, beag go leor (240k focal) i gcomparáid le corpais labhartha do theangacha eile amhail an Béarla, m.sh. Corpas Náisiúnta na Breataine ina bhfuil 10 milliún focal de shonraí labhartha tras-scríofa, arb ionann é agus 10% den chorpas 100 milliún focal, agus Corpas Idirnáisiúnta an Bhéarla , ar corpais iad uile ina bhfuil 600k focal de Bhéarla labhartha.

Forbairtí Riachtanacha Amach Anseo
- Cé go gcuireann an corpas labhartha Comhrá bunús maith ar fáil, tá sé neamhiomlán i mórán de na réimsí a shonraítear sa dearadh corpais tosaigh. Is gá an corpas a leathnú go suntasach, go háirithe trí réimsí sainiúla úsáide teanga a chur leis, m.sh. suíomhanna gnó, dioscúrsa ranga, seomraí cúirte, clinicí dochtúirí, ospidéil, clubanna spóirt, ionaid fóillíochta, salóin ghruaige, ollmhargaí srl.
- I gcás an chorpais labhartha, ní mór aird ar leith a thabhairt ar chainníocht na sonraí atá le taifeadadh agus le tras-scríobh, líon na ndaoine a bheidh le taifeadadh, a gcúlra teanga, a n-aois, a gcanúint, anuas ar a dhoimhne atá anótáil déanta ar an gcorpas. I measc na bhfachtóirí breise atá le cur san áireamh tá caighdeán agus cineál an taifeadta, an trealamh beacht taifeadta agus an timpeallacht taifeadta, agus cineál na n-ábhar atá le taifeadadh.
- Is féidir sonraí tras-scríofa reatha (m.sh. podchraoltaí RnaG) a chur san áireamh ach an tras-scríbhinn a ailíniú leis an gcomhad fuaime má tá sé ar fáil.
- Is acmhainn luachmhar é corpas labhartha do mhúineadh agus d'fhoghlaim teanga. Chun an tairbhe is mó a bhaint as, ba cheart na tras-scríbhinní fuaime a chatagóiriú de réir catagóirí éagsúla oideolaíocha, amhail ábhar, luas agus castacht na cainte, soiléireacht na cainte, srl.

2.4 An Scéalaí

Corpas foghlaimeoirí é seo atá á bhailiú mar chuid d'ardán foghlama ríomhchuidithe teangacha cliste atá á fhorbairt faoi láthair. Is é an sprioc atá ag an Scéalaí ná cruthú ábhair scríofa a spreagadh ó fhoghlaimeoirí, agus ag an am céanna iad a nochtadh do leaganacha labhartha dá dtéacs féin (ag úsáid guthanna sintéiseacha ABAIR) agus aschur cainte a mhealladh. Mar chuid den ardán, cuireann na foghlaimeoirí eolas ar fáil faoin gcúlra sochtheangeolaíoch, mar aon le féinmheasúnú ar a leibhéal teanga, rud a mheaitseáiltear ina dhiaidh sin lena n-aschur scríofa agus labhartha. Tá na sonraí uile anaithnid.

Forbairtí Riachtanacha Amach Anseo
Is corpais foghlaimeoirí an-nua iad seo agus de réir mar a fhásfaidh siad, beidh gá le huirlisí agus creataí a fhorbairt ionas gur féidir anailís a dhéanamh orthu chun bonn eolais a chur faoin taighde ar shealbhú an dara teanga agus faoi na tosaíochtaí d'fhorbairt na foghlama ríomhchuidithe teanga amach anseo, lena n-áirítear leaganacha nua den Scéalaí amach anseo.

2.5 Corpas Aitheanta Cainte ABAIR MíleGlór

Tá corpas MíleGlór á bhailiú faoi láthair chun córais aitheanta cainte Gaeilge a fhorbairt. Tá cur chuige sluafhoinsithe spriocdhírithe in úsáid agus ionchur ag an pobal tríd an bhfón agus tríd an ríomhaire ag https://www.abair.tcd.ie/mileglor. Tá taifeadtaí allamuigh ar siúl i láithreacha Gaeltachta freisin. Tá samplaí beaga cainte á mbailiú ó líon mór cainteoirí, a chlúdaíonn gach canúint, gach aois, gach leibhéal. Tá an obair seo á comhordú leis an mbailiúchán corpas do shintéis urlabhra (Féach §2.1, 2.2 thuas). Breactar síos faisnéis shochtheangeolaíoch ábhartha roimh sheisiún taifeadta agus déantar an fhaisnéis sin a ailíniú le cód aitheantais a ghabhann leis an taifeadadh.

San áireamh i gcorpas MíleGlór chomh maith tá a lán sonraí cainte a bhí ann cheana, mar aon le téacs (m.sh., ó fhoinsí meán cumarsáide).

Forbairtí Riachtanacha Amach Anseo
- Taifeadtaí fairsinge, ag cur leis an gclúdach reatha. Is féidir modhanna sluafhoinsithe a úsáid chun an bailiú seo a dhéanamh (CITE Welsh Padua App).
- Bailiú ábhar eile atá ann cheana is féidir a úsáid.
- Glanadh an chorpais (profeisteacht m.sh., le hábhar neamhábhartha a bhaint, stadanna sa chaint a lipéadú srl.)
- Ardán cobhsaí chun na taifeadtaí ar líne ó fhoinse ar bith a bhainistiú. D'fhéadfadh sé go n-áireofaí leis seo aip shainoiriúnaithe iOS agus Android le haghaidh taifeadadh, a bheadh éasca le húsáid.
- Freastalaí agus bainistiú lena chinntiú go ndéantar na sonraí a óstáil, a chosaint agus a úsáid go héifeachtach i dtógáil córas aitheanta cainte (féach Caibidil 9), agus feidhmchláir.
- Tá gá le bailiú forleathan samplaí cainte le haghaidh córais chruinne aitheanta cainte. Braitheann córais uathoibríocha fotheidealaithe ar thaifeadtaí cainte agus tras-scríbhinní ailínithe a bheith ar fáil.

2.6 Corpas Foghlaimeoirí Gaeilge Scríofa agus Labhartha TEG

Is éard atá sa chorpas faoi láthair 200,000 focal de shonraí labhartha agus scríofa foghlaimeoirí, ó scrúduithe Theastas Eorpach na Gaeilge, ag leibhéil meánleibhéil agus ardleibhéil. Féach cuid 1.10 Corpas Foghlaimeoirí Gaeilge Scríofa agus Labhartha TEG le tuilleadh sonraí a fháil.

2.7 Corpas Inchomparáide Idirnáisiúnta: (ICCI) Gaeilge Labhartha agus Scríofa

Is é aidhm an tionscadail Corpas Inchomparáide Idirnáisiúnta (CII) ná sraith corpas inchomparáide a chruthú i roinnt teangacha lena n-áirítear an Ghaeilge (áirítear Seicis, Béarla, Fionlainnis, Fraincis, Gearmáinis, Iodáilis, Ioruais, Polainnis, Slóvaicis agus Sualainnis i measc na dteangacha eile). Beidh 1 mhilliún focal i ngach corpas, is é sin 40% de theanga scríofa agus 60% de theanga labhartha. I gcorpas CII baintear úsáid as dearadh corpais Chorpas Idirnáisiúnta an Bhéarla chun an inchomparáideacht is mó is féidir a dhéanamh leis na corpais reatha don Bhéarla (Béarla na hÉireann san áireamh) agus athúsáidfear acmhainní teanga atá ann cheana nuair is féidir.

Gné thábhachtach uathúil de gach corpas CII is ea gur páirt shuntasach de an chaint spontáineach, comhdhéanta de thuairim agus 600,000 focal (60% den chorpas). Ach sonraí labhartha den chineál sin thar 15 chás dioscúrsa nó níos mó a chur ar fáil le haghaidh anailís chodarsnach i measc roinnt teangacha, beifear in ann comparáidí tras-teangeolaíochta corpasbhunaithe ar an teanga labhartha a dhéanamh, rud a bhfuil géarghá leis agus nach ndearnadh riamh roimhe [32]. Mar aon le sonraí cothromaithe ar fud na réimeanna scríofa, beidh CII ina acmhainn fhíorluachmhar don mheaisínfhoghlaim, mar shampla meaisínaistriúchán [33], don fhoclóireacht chorpasbhunaithe agus don taighde teangeolaíoch.

Forbairtí Riachtanacha Amach Anseo
- Measúnú a dhéanamh ar infhaighteacht acmhainní oiriúnacha scríofa agus labhartha atá le húsáid in CII. Is dócha go mbeidh go leor acmhainní scríofa ar fáil don chomhpháirt scríofa (400k focal). Mar sin féin, meastar go gcaithfear an chuid is mó den pháirt labhartha den chorpas (600k focal) a thaifeadadh, a thrascríobh agus a anótáil. Cé go bhfuil go leor closchomhaid chraolta (podchraoltaí) agus tras-scríbhinní ar fáil, m.sh. sa chorpas cainte Comhrá (féach cuid 2.3 srl), ní hionann an réimse úsáide sin ach 10% de dhearadh an chorpais labhartha. I gcodarsnacht leis sin, is fíorbheagán sonraí cainte spontáineacha atá ann do na cásanna dioscúrsa eile, arb ionann iad agus 90% den chuid labhartha de CII.