



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Staid Reatha an Mheaisínáistriúcháin Gaeilge
Author(s)	Lane, Caoilfhionn
Publication Date	2019
Publication Information	Lane, Caoilfhionn. (2019). Staid Reatha an Mheaisínáistriúcháin Gaeilge. Léann Teanga: An Reiviú, https://doi.org/10.13025/f843-5w83
Publisher	Acadamh na hOllscolaíochta Gaeilge, Ollscoil na hÉireann, Gaillimh
Link to publisher's version	https://doi.org/10.13025/f843-5w83
Item record	http://hdl.handle.net/10379/17494
DOI	http://dx.doi.org/10.13025/f843-5w83

Downloaded 2024-05-14T17:15:29Z

Some rights reserved. For more information, please see the item record link above.





Staid Reatha an Mheaisínastriúcháin Gaeilge

Caoilfhionn Lane

Réamhrá

Idir 2015 agus 2017, mar chuid dá rannóg grinn, d'fhoilsigh tuairisc.ie sraith alt le samplaí den drochaistriúchán Gaeilge i gcáipéisí agus in áiteanna poiblí ar fud na tíre. Sa tsraith sin, *Amú san aistriúchán*, foilsíodh samplaí ar nós 'Leabh cheantar atá ag athrú' mar aistriúchán ón mBéarla 'baby changing zone'.¹

Dá mbeadh aistritheoir gairmiúil ag iarraidh Gaeilge a chur ar an mBéarla sin, d'fhéadfaidís leas a bhaint as foclóir.ie, ina bhfuil *seomra gléasta páiste*, *seomra athraithe páiste*² tugtha, agus tá *ionad friothála ar naíonáin* ar fáil ó tearma.ie³. Is léir, mar sin, nach ndeachaigh aistritheoir gairmiúil, coinsiasach i ngleic le 'baby changing zone'. Go deimhin, drochsheans gur cainteoir Gaeilge a bhuaíl faoin aistriúchán sin.

Ach cár bh as a dtáinig an leagan Gaeilge mar sin? Seans láidir gur meaisínastriúchán a bhí in úsáid ag pé duine a bhí i mbun aistriúcháin. Tá dul chun cinn suntasach déanta i réimse an mheaisínastriúcháin le blianta beaga anuas (Bojar *et al.*, 2016), ach caithfear a bheith cúramach faoin gcaoi a gcuirtear an teicneolaíocht i bhfeidhm. Is iomaí sampla atá ar fáil i nGaeilge de drochaistriúcháin (nó de 'phraistriúcháin', mar a thugtar orthu sna meáin), sa tsraith a d'fhoilsigh tuairisc.ie agus i bhfoinsí eile, agus is léir go n-eascraíonn go leor de na samplaí sin as mí-úsáid an mheaisínastriúcháin.

1 <https://tuairisc.ie/amu-san-aistriuchan/>

2 [https://www.focloir.ie/en/dictionary/ei/changing+room#changing-room__2\(\)](https://www.focloir.ie/en/dictionary/ei/changing+room#changing-room__2())

3 <http://www.tearma.ie/Search.aspx?term=baby+changing>

Mar sin, is ábhar conspóideach i gcónaí é an t-aistriúchán in áiteanna poiblí (agus úsáid an mheaisínáistriúcháin sa chomhthéacs sin), go háirithe i gcás mionteangacha. Mar shampla, i mí Aibreáin 2019, chinn an Coimisinéir Teanga gur sháraigh Comhairle Contae Chorcaí a gcuid dualgas nuair a d'úsáid siad *Google Translate* chun a suíomh gréasáin a aistriú go Gaeilge agus bhí an t-aistriúchán lochtach⁴.

San alt seo, déanfar athbhreithniú ar an teicneolaíocht seo i gcomhthéacs na Gaeilge agus pléifear na forbairtí is déanaí sa réimse. Breathnófar freisin ar na gnéithe comhchoiteann a théann i bhfeidhm ar úsáideoirí agus ar fhorbróirí an mheaisínáistriúcháin: caighdeán an aschuir ó chórais mheaisínáistriúcháin Gaeilge, agus úsáid chuí agus mí-úsáid na teicneolaíochta. Pléifear freisin na difríochtaí idir córas ilfheidhmeach ar nós *Google Translate*, agus córas le sainfeidhm, mar shampla córas meaisínáistriúcháin a bhaineann le teanga an rialtais amháin.

Úsáid Chuí an Mheaisínáistriúcháin

Tá an-tóir ar *Google Translate* mar uirlis mheaisínáistriúcháin agus in 2016 rinne *Google* comóradh ar dheich mbliana ó cuireadh an uirlis ar fáil, uirlis a bhí ag freastal ar 500 milliún duine gach lá ag deireadh na tréimhse sin⁵. Tá na bogearraí seo saor in aisce ar líne, agus ag úsáid an chomhéadain is féidir leis an úsáideoir leagan aistrithe de théacs a nochtadh, i gcibé rogha teanga atá ag an úsáideoir. Tá *Google Translate* in ann os cionn 100 teanga a láimhseáil, an Ghaeilge san áireamh, ach ní gá go mbeadh an caighdeán aistriúcháin céanna ag dul le gach teanga, mar go láimhseáiltear na teangacha ar bhealaí éagsúla⁶.

Cé go dtéann úsáideoirí difriúla i ngleic le teicneolaíocht mheaisínáistriúcháin ar nós *Google Translate* ar bhealaí éagsúla, ag brath ar a gcuid riachtanas agus leibhéal cumais i dteangacha éagsúla, is féidir a rá go bhfuil dhá mhórghrúpa úsáideoirí ag *Google Translate* – úsáideoirí gairmiúla agus úsáideoirí nach úsáideoirí gairmiúla iad, úsáideoirí ócáideacha ina measc siúd.

Is minic a bhíonn an **t-úsáideoir ócáideach** ag iarraidh eolas simplí a aimsiú, nó a chur in iúl. Ag úsáid comhéadain de chuid *Google Translate*, is féidir leis an úsáideoir ócáideach éirim an scéil a bhaint amach i dteanga nach bhfuil ar eolas aige nó aici. Mar shampla, d'fhéadfadh cainteoir aonteach Béarla teach saoire a chur in áirithe i nGearmáinis gan stró ag úsáid *Google Translate*. Sa chás seo, tá ról an úsáideora agus leas an mheaisínáistriúcháin soiléir: tá an t-úsáideoir ócáideach ag iarraidh eolas simplí a fháil amach, agus is leor aistriúchán garbh chun an sprioc sin a bhaint amach. Tá a mhalairt fíor freisin, mar bhainfeadh an cainteoir Gearmáinise ar bheagán Béarla an luach céanna amach as aistriúchán garbh ar nasc Béarla.

4 <https://tuairisc.ie/cinneadh-deanta-ag-an-gcoimisineir-teanga-faoi-chearta-teanga-ar-twitter-agus-usaid-google-translate/>

5 <https://blog.google/products/translate/ten-years-of-google-translate/>

6 <https://cloud.google.com/translate/docs/languages>

Cé go gceapfaí go luíonn sé seo le réasún, ní gá gurb amhlaidh atá sé seo le gach teanga. Ar nós fhormhór na dteangacha coitianta, tá *Google Translate* in ann Gaeilge a aistriú go teangacha eile agus *vice versa*. Ach, go minic, ní bhaineann pobal na Gaeilge leas as, mar go bhfuil Béarla acu go léir. Is iad úsáideoirí gan Ghaeilge a úsáideann *Google Translate* (Gaeilge) den chuid is mó. Tugtar úsáid aontaobhach air seo. Is nósmaireacht coitianta go maith é seo i gcás na mionteanga.

Bhí an méid seo le rá ag (Měchura, 2017) faoin úsáid ócáideach seo i gcás an mheaisínáistriúcháin:

... meaisínáistriúchán le haghaidh tuisceana, chun téacs atá i dteanga nach bhfuil ag an léitheoir a dhéanamh sothuigthe dó nó di. Is féidir an meaisínáistriúchán a úsáid chun aistriúchán garbh a chruthú agus, cé go mbeidh sé beagáinín ciotach agus mícheart, tabharfaidh sé éirim an téacs don léitheoir. (Měchura, 2017: 67)

Tá riachtanais éagsúla ag **úsáideoirí gairmiúla**. De ghnáth is aistritheoirí proifisiúnta iad seo, ach d'fhéadfaí gur iar-eagarthóirí iad freisin. Déanann iar-eagarthóirí eagarthóireacht ar an gcéad dréacht de phíosa aistriúcháin. Ar na huirlisí atá ina gcabhair don aistritheoir agus don iar-eagarthóir tá: foclóirí, bunachair téarmaíochta, agus an t-aistriúchán ríomhchuidithe. Ach, baineann iar-eagarthóirí leas as an meaisínáistriúchán chun an chéad dréacht de phíosa aistriúcháin a chruthú. Go deimhin, is as teacht réimse an mheaisínáistriúcháin a d'eascair ról an iar-eagarthóra (Church & Hovy, 1993).

Thug Offersgaard *et al.* (2008) achoimre ar na scileanna atá ag teastáil go hidéalach ón iar-eagarthóir nó an t-aistritheoir atá ag baint úsáid as an meaisínáistriúchán don chéad dréacht:

A good post-editor knows the domain; A good post-editor has very good skills in his/her native language; A good post-editor can implement style guides and consistent terminology; A good post-editor is an experienced proof-reader. He/she can make swift decisions on 'good – no good' in order to be able to discard machine-translated text strings that are not worthwhile post-editing, but need to be translated from scratch.

I gcás an aistriúcháin ríomhchuidithe faigheann an t-aistritheoir méadú ar tháirgiúlacht ag úsáid sainbhogearraí ar nós *SDL Trados*⁷ nó *MemoQ*⁸. Ach cén chaoi a bhfuil caighdeán an aschuir ón meaisínáistriúchán i comparáid leis an aistriúchán ríomhchuidithe? Braitheann sé sin ar an gcineál algartaim atá sa chóras meaisínáistriúcháin, agus ar an gcóras é féin. Ar

7 <https://www.sdltrados.com/>

8 <https://www.memoq.com/en/>

chúiseanna éagsúla, is féidir dhá chóras meaisínáistriúcháin bunaithe ar an algartam céanna a chur i gcomparáid agus ní gá go mbeadh an caighdeán céanna acu. Míníonn Mëchura (2017) an comhthéacs ina bhfuil an meaisínáistriúchán inmholta agus ina chabhair don aistritheoir:

...mar chuid de phróiseas an aistriúcháin ina ndéantar meaisínáistriúchán ar dtús, ansin déanfaidh aistritheoir nó profléitheoir daonna iar-eagarthóireacht air. Is amhlaidh a bhíonn sé níos tapúla, uaireanta, téacs a aistriú le meaisínáistriúchán go gceartóidh duine na botúin ansin, seachas duine á aistriú ón tús. Bíonn próiseas mar seo ar siúl ag eagraíochtaí a mbíonn orthu líon mór téacasanna a aistriú arís agus arís eile... (Mëchura, 2017: 68)

Ach tugann sé foláireamh don té a bheadh ag iarraidh meaisínáistriúchán a úsáid chun tabhairt faoi pháosa a aistriú:

Níl sé inmholta, mar shampla, an meaisínáistriúchán a úsáid chun suíomh idirlín nó bróisiúr a aistriú go teanga éigin nach bhfuil ar do thoil agat agus an t-aistriúchán a fhoilsiú go poiblí. (Mëchura, 2017: 68)

Is dócha nach bhfuil an chomhairle seo á leanúint i roinnt cásanna de na ‘praistriúcháin’ ar chomharthaí poiblí atá fós ag teacht aníos sna meáin. Go deimhin, d’fhéadfaí a rá gur ar an úsáideoir atá an locht, ní ar an teicneolaíocht. Bheifí ag súil go mbeadh úsáideoir oilte nó úsáideoir gairmiúil ag aistriú comhartha poiblí, ach seans gur chomhlíon úsáideoir ócáideach an ról sa chás sin, ní dhearnadh machnamh ar an tasc, agus baineadh mí-úsáid as inneall meaisínáistriúcháin.

Samhlacha Meaisínáistriúcháin

Tá bealaí éagsúla ann chun córas meaisínáistriúcháin (MT) a fhorbairt: meaisínáistriúchán bunaithe ar rialacha nó **meaisínáistriúchán riailbhunaithe (RBMT)**, an **meaisínáistriúchán staitistiúil (SMT)**, an **meaisínáistriúchán néarach (NMT)** agus an **meaisínáistriúchán hibrideach**. Déanann na téarmaí SMT⁹ nó NMT¹⁰ cur síos ar an gcineál algartaim, agus tugtar samhail ar an táirge a chruthaíonn algartam ar leith. Tá buntáistí agus míbhuntáistí ag baint le gach cineál córais, agus bíonn caighdeán an aschuir éagsúil ag brath ar an tsamhail a úsáidtear chomh maith leis an mbuntéacs atá le haistriú. Tá córais aistriúcháin éagsúla bunaithe ar shaineolas teanga agus scríobhann forbróirí rialacha aistriúcháin dóibh. Tá córais eile bunaithe ar chorpas téacs a bheith ar fáil agus léiriú staitistiúla de rialacha teangeolaíochta

9 *Statistical Machine Translation*

10 *Neural Machine Translation*

a bhaint den téacs le halgartaim. Tá caighdeáin na gcóras seo taobh leis an saineolas sa chéad chás nó le méad an chorpais sa dara cás. Déantar cur síos ar na samhlacha seo thíos. Tabhair faoi deara, go mb'fhéidir go bhfuil córas meaisínáistriúcháin ag úsáid níos mó ná samhail amháin. Mar shampla, i gcás *Google Translate*, is minic go mbraitheann an tsamhail atá in úsáid ar na teangacha atá i gceist.¹¹

Meaisínáistriúchán riailbhunaithe

Bhí na chéad chórais mheaisínáistriúcháin bunaithe ar rialacha teangeolaíochta, agus tá an cur chuige seo fós in úsáid i bhforbairt córais mheaisínáistriúcháin. Ach bíonn go leor laincísí ag baint leis an gcur chuige sin, go príomha mar tá saineolas sa dá theanga ag teastáil ón bhforbróir (an ríomheolaí nó an t-innealtóir bogearraí) ar an dá theanga atá i gceist. Caithfidh córais **riailbhunaithe (RBMT)** cruinne a bheith in ann rialacha casta teanga, gach gné den ghramadach, agus eisceachtaí a láimhseáil. Buntáiste a bhaineann le córais mar seo na nach bhfuil corpas mór téacs de dhíth chun iad a fhorbairt, agus feileann sé seo do roinnt teangacha (Forcada *et al.*, 2011).

Meaisínáistriúchán staitistiúil

Le gairid, is iad na córais **mheaisínáistriúcháin staitistiúla (SMT)** is iomadúla in úsáid. Tá córas SMT bunaithe ar **chorpas comhthreomhar**. Is éard is corpas comhthreomhar ann ná bailiúchán de phéirí abairtí sa dá theanga, agus a bhfuil an bhrí chéanna ag gach péire acu.

Go ginearálta, is féidir a rá go bhfuil patrúin chomónta le haithint sa dá theanga nuair atá an t-ábhar céanna faoi chaibidil. Tá sé seo le feiceáil go soiléir i gcorpas comhthreomhar. Is féidir le córas meaisínáistriúcháin samhail staitistiúil den ghaol idir an dá theanga sa chorpas comhthreomhar seo a chruthú. Déanann Měchura (2017) cur síos ar an bpróiseas seo:

As sin, cruthóidh an ríomhaire samhail staitistiúil den dá theanga agus den ghaol eatarthu. Go bunúsach, tá an ríomhaire ag tabhairt faoi phróiseas foghlama, é ag obair as a stuaim féin faoi stiúir ag algartam staitistiúil, gan d'ionchur aige ach an fhianaise atá le fáil sa chorpas agus gan tuiscint aige ar bhrí na n-abairtí. (Měchura, 2017: 66)

11 <https://cloud.google.com/translate/docs/languages>

Ach is aistritheoirí daonna a chruthaíonn na corpais chomhthreomhara atá riachtanach d'fhorbairt córas meaisínáistriúcháin. Míníonn Kenny (2012) an tábhacht a bhaineann leis na haistriúcháin dhaonna sin mar shlat tomhais:

SMT thus relies on human translation for its training data, but it also relies on human translation for its legitimacy: the reason developers of SMT systems use parallel corpora to train their systems is because such corpora are assumed to contain good answers to translation problems; and they are assumed to contain good answers precisely because they contain translations performed by human beings. (Kenny, 2012)

Chomh maith leis an gcorpas comhthreomhar, úsáidtear **corpais aonteangach** i gcórais SMT chun mionshamhlacha staitistiúla a dhéanamh de na patrúin a bhaineann le teanga amháin gan tagairt don teanga eile. Sa chaoi sin, tá níos mó na samhail amháin staitistiúil ag an gcóras, an ceann a bhaineann leis na bealaí a gcomhfhreagraíonn na teangacha lena chéile, agus samhlacha aonteangacha freisin a bhaineann le rialacha gramadaí na dteangacha.

Córas meaisínáistriúcháin hibrideach

Is féidir **córas meaisínáistriúcháin hibrideach** a chruthú freisin. Mar shampla, is féidir dréacht a chruthú le córas SMT mar chéad chéim agus ansin rialacha gramadaí breise a chur i bhfeidhm ar an dréacht uaidh ionas gur féidir an cruinneas a fheabhsú (Dowling *et al.*, 2016).

Meaisínáistriúchán néarach

Le cúpla bliain anuas, tá samhail meaisínáistriúcháin nua in úsáid: an **meaisínáistriúchán néarach** (NMT ó *neural machine translation* i mBéarla) (Bojar *et al.*, 2016). Tá córas NMT bunaithe ar líonraí néaracha saorga córais atá iad féin bunaithe go scaoilte ar na líonraí néaracha in inchinn dhaonna. Dá thoradh seo bíonn go leor algartam in úsáid in aon chóras amháin.

De ghnáth bíonn caighdeán an aschuir ó chóras NMT níos fearr ná na córais eile, ach tá na riachtanais ó thaobh méid téacs de fós bailí (Bojar *et al.*, 2016). Mar sin, is féidir torthaí maithe a bhaint amach sna mórtheangacha le NMT. Ag deireadh 2016, thug *Google* an NMT isteach mar cheann de na samhlacha atá in úsáid ag *Google Translate*. Ar dtús, bhí *Google* NMT ar fáil d'ocht teanga (Wu *et al.*, 2016) agus tugadh isteach an NMT do go leor teangacha eile (an Ghaeilge san áireamh) de réir a chéile. Ionas gur féidir córas NMT a fhorbairt, tá corpais comhthreomhar ag teastáil, ach is féidir córas NMT a fhorbairt gan téacs aonteangach.

Caighdeán Aistriúcháin ó Chórais Meaisínáistriúcháin

Tá difríochtaí idir na samhalacha éagsúla ó thaobh caighdeáin de. Mar shampla, de ghnáth bíonn na haistriúcháin a eascraíonn ó NMT níos inléite i gcomparáid leis na haistriúcháin ó SMT (Bojar *et al.*, 2016). Ach braitheann an difríocht sa chaighdeán idir dhá chóras meaisínáistriúcháin freisin ar na mionsonraí a bhaineann leis an dá chóras agus ar ghnéithe eile. Chomh maith leis an tsamhail atá in úsáid i gcóras meaisínáistriúcháin, tá tionchar ag téacs an chorpais chomhthreomhair agus an corpas aonteangach ar chaighdeán an aschuir.

Méid an Chorpais

Ní féidir samhail ar ardchaighdeán a chruthú gan téacs cruinn sa chorpas comhthreomhar agus, i gcás SMT, don chorpas aonteangach. Deirtear go minic i réimse na ríomheolaíochta (agus réimsí eile atá ag brath ar chruinneas sonraí): ‘dramhaíl isteach, dramhaíl amach’ (Kim, Huang & Emery, 2016). Mura bhfuil corpas téacs cruinn ag dul isteach, ní féidir dul i muinín an téacs a thagann amach. Ní hamháin go gcaithfidh an téacs a bheith cruinn, ach tá go leor téacs (na milliúin focal) ag teastáil chun córas athléimneach a chruthú.

Ní furasta líon mór téacs atá i dteangacha éagsúla a aimsiú, agus bíonn an-tóir ar na corpais chomhthreomhara atá amuigh ansin, go háirithe corpais atá i dteangacha nach mórtheangacha domhanda iad. Is foinse mhór de chorpais chomhthreomhara é an tAontas Eorpach (AE), ach cé go bhfuil an Ghaeilge ar cheann de theangacha oifigiúla an AE ó 2007, mar gheall ar mhaolú na Gaeilge san AE, ní gá go n-aistrítear go Gaeilge gach aon cháipéis a fhoilsíonn institiúidí de chuid an AE. Dá thoradh sin, bíonn na corpais chomhthreomhara beag i gcomparáid le teangacha eile an AE. Beidh maolú na Gaeilge san AE i bhfeidhm go dtí deireadh na bliana 2021.

Tá staid na Gaeilge i gcomparáid leis na teangacha oifigiúla eile soiléir i gcás *DGT-TM: Directorate General for Translation*, corpas comhthreomhar atá ar fáil ón Aontas Eorpach:

The first version of DGT-TM was released in the year 2007, including EU documents up to the year 2006. There have been three updates since (releases 2011, 2012 and 2013) and it is planned to release new data every year. The data up to the year 2013 includes 23 languages (all official EU languages except Croatian), but the 23rd language, Irish, is as usual much under-represented. The next release of DGT-TM (release 2014, including the data for the year 2013) is expected to include 30,000 Irish and about 200,000 Croatian Translation Units (TUs). (Steinberger et al., 2014)

Níl an deacracht chéanna le corpas aonteangach: is féidir corpas nua a chruthú ó aon téacs Gaeilge fad is go bhfuil sé ar fáil i bhfoirm atá inléite ag ríomhairí. Níl aon laincis téacs Gaeilge ó fhoinsé oscailte (mar shampla, téacs ón Vicipéid), i bhfoirm atá inléite ag ríomhaire,

a íoslódáil chun é a úsáid. Is féidir freisin comhéadan feidhmchláir (API) nó cnuaschoipeáil ón ngréasán a úsáid chun é a íoslódáil go huathoibríoch. Fós féin, níl an méid céanna téacs oscailte ar fáil i nGaeilge i gcomparáid leis na mórtheangacha.

Réimse Teanga an Chorpais

Chomh maith le líon na bhfocal i gcorpas éigin, bíonn an friotal sa chorpas an-tábhachtach mar ionchur do chóras meaisínáistriúcháin. Bíonn tionchar freisin ag réimse teanga an chorpais ar an gcóras aistriúcháin a eascraíonn as. Má tá córas ilfheidhmeach nó uileghabhálach ag teastáil, bíonn téacs ó réimsí teanga éagsúla ag teastáil. Ar an lámh eile, má tá córas le bheith dírithe ar úsáideoirí agus iad ag plé le réimse teanga amháin – aistritheoirí ar cáipéisí leighis, cuir i gcás – bíonn cáipéisí leighis ag teastáil mar ionchur, mar go mbíonn na cáipéisí sin lán le samplaí den téarmaíocht agus den stíl chuí. Ar an gcaoi chéanna, má tá tvuít nó giolc le haistriú, is iad am-línte Twitter an corpais is fearr le húsáid.

Léiríonn na samplaí seo cuid bheag den éagsúlacht teanga, ach taispeánann siad dúshlán amháin a bhaineann le forbairt córas ilfheidhmeach. Go ginearálta, bíonn sé deacair aistriú gan leid faoin gcomhthéacs, agus tá an rud céanna amhlaidh don mheaisínáistriúchán. Is féidir dul i ngleic leis an dúshlán sin le córas atá dírithe ar réimsí teanga ar leith (leigheas, stair, eolaíocht, srl.) ach is é an toradh air sin go mbeadh córais iomadúla le forbairt do na réimsí ar leith teanga, cur chuige atá níos costasaí agus níos fadálaí. Dar le Offersgaard *et al.* (2008) is comhréiteach an réiteach is fearr:

...in principle it would be preferable to train an SMT system on texts with almost identical writing styles and within the same subject domain. On the other hand, for practical and financial reasons, it would be desirable that the SMT system had a broad coverage being usable for different text types without a negative impact on the translation quality. So, the solution is a compromise.

Measúnú Caighdeán an Mheaisínáistriúcháin

Tá forbróirí agus iar-eagarthóirí araon ag súil le bunchaighdeán ar leith ó chóras meaisínáistriúcháin, is é sin go bhfuil an t-aistriúchán inléite. Ach bíonn tosaíochtaí breise ag aistritheoirí. Níl siad ag iarraidh an iomarca ama a chaitheamh i mbun iar-eagarthóireachta agus tá siad ag iarraidh comhéadan iar-eagarthóireachta atá inúsáidte (Offersgaard *et al.*, 2008).

Úsáideann aistritheoirí agus forbróirí modhanna difriúla chun caighdeán an aschuir a mheas (Sun, 2010). Go minic, úsáideann forbróirí na gcóras aistriúcháin slata tomhais ar leith chun caighdeán na n-aistriúcháin a mheas agus chun córais a chur i gcomparáid lena chéile (Papineni *et al.*, 2002; Turian, Shen and Melamed, 2003; Snover *et al.*, 2006). Déanann na slata tomhais is coitianta luacháil ar chaighdeán an chórais bunaithe ar rialacha agus critéir

ar leith, agus is i bhfoirm uimhreach atá an scór deiridh. Cé go n-úsáideann forbróirí na scóir seo chun córais a chur i gcomparáid lena chéile, níl cuid acu an-iomasach d'aistritheoirí agus ní gá go mbeadh an luach céanna ag aistritheoirí ar an scór seo agus iad ag meas chaighdeán an aistriúcháin.

Mar a léiríonn Sun (2010): '*... as humans are the end-users of MT, human judgement is ultimately the benchmark to assess the usefulness of automatic metrics. How good an automatic metric is depends on its correlation with human evaluation.*' Uaireanta, úsáidtear slata tomhais atá níos iomasaí d'aistritheoirí, mar shampla an slat tomhais TER (*Translation Edit Rate*): '*Translation Edit Rate (TER) measures the amount of editing that a human would have to perform to change a system output so it exactly matches a reference translation.*' (Snover et al., 2006)

Is féidir, chomh maith, comparáid dhíreach a dhéanamh idir córas meaisínáistriúcháin agus uirlis aistriúcháin eile – mar shampla, bogearraí cuimhne aistriúcháin (TM). Phléigh Moorkens & Way (2016) staidéar amháin ar an aistriúchán idir Spáinnis agus Béarla i réimse teicniúil ar leith. D'aistrigh iar-eagarthóirí sleachta téacs ag úsáid córas meaisínáistriúcháin nó bogearraí cuimhne aistriúcháin. Dar leis an staidéar sin, chaith na hiar-eagarthóirí an méid céanna ama ag déanamh eagarthóireachta ar aschur an mheaisínáistriúcháin agus a chaith siad nuair a d'úsáid siad na bogearraí cuimhne aistriúcháin le meaitseáil neamhbheacht (*fuzzy match*) de 85 – 94% ann.

D'ardaigh Moorkens & Way (2016) ceist lárnach faoi chaighdeán an mheaisínáistriúcháin i gcomparáid le TM:

Is comparative acceptability of TM over MT predicated on the user's ability to optimise the precision and usefulness of match suggestions by setting a minimum match threshold? We contend that the answer to this question is yes, and that:

1. *MT would be considered more acceptable to users if only those matches that required relatively small amounts of editing were presented to post-editors.*
2. *TM would be less acceptable to users if matches that required large amounts of editing were presented to translators. (Moorkens & Way, 2016)*

Chomh maith leis na slata tomhais sin, is féidir caighdeán meaisínáistriúcháin a luacháil de réir scéimre ar leith. Léann measúnóirí abairtí ó chóras meaisínáistriúcháin agus déanann siad measúnú suibíachtúil de réir scéimre ar nós *Ideal/Acceptable/Possibly acceptable/Unacceptable* (Coughlin, 2003).

Athbhreithniú ar an litríocht: Meaisínaistriúchán na Gaeilge

Tá go leor athruithe le cúpla bliain anuas i réimse an mheaisínaistriúcháin agus baineann na hathruithe sin le meaisínaistriúchán na Gaeilge freisin. Tháinig feabhas ar chaighdeán na n-aistriúcháin ó *Google Translate* nuair a d'athraigh an comhlacht an tsamhail aistriúcháin ó SMT go NMT (Wu *et al.*, 2016). Cuireadh an tsamhail NMT ar fáil chun aistriúcháin a dhéanamh idir Béarla agus dornán mórtheangacha in 2016. De réir a chéile, mhéadaigh Google líon na dteangacha leis an tsamhail NMT, agus anois tá an Ghaeilge ina measc.

Chomh maith le *Google Translate*, tá córais mheaisínaistriúcháin eile a aistríonn go Gaeilge. Ar cheann de na córais is luaithe acu tá córas Scannell (2006). Forbraíodh an córas riailbhunaithe sin ionas gur féidir aistriú idir an Ghaeilge agus Gàidhlig na hAlban, agus d'éirigh go maith leis an gcóras sin mar gheall ar an gcosúlacht idir an dá theanga. Níor forbraíodh córas Béarla–Gaeilge, áfach, go dtí níos déanaí, agus is córais SMT agus NMT iad.

De réir Acht na dTeangacha Oifigiúla 2003, bíonn dualgais teanga ar anS stát de réir rialachán, scéimeanna, reachtaíochta agus eile. Chun cabhrú leis an éileamh ar aistriúcháin ar cháipéisí riaracháin go Gaeilge mar gheall ar an Acht, d'fhorbair Dowling *et al.*, (2015) córas SMT dírithe ar aistritheoirí atá ag obair san earnáil riaracháin phoiblí. Cuireadh feabhas ar an gcóras sin leis an iar-eagarthóireacht uathoibríoch (Dowling *et al.*, 2016) agus tá an córas hibrideach sin in úsáid sa Roinn Cultúir, Oidhreacht agus Gaeltachta. Chomh maith leis an gcóras hibrideach sin, a bhaineann go sainiúil le réimse teanga amháin, d'fhorbair Arčan *et al.* (2016) córas SMT ilfheidhmeach a aistríonn ó Bhéarla go Gaeilge agus a mhalairt treo.

Córas	Scannell (2006)	Dowling et al. (2015, 2016)	Arčan et al. (2016)	Dowling et al. (2018)	Google Translate
Samhail	RBMT	SMT	SMT	SMT, NMT	SMT, NMT
Bunteanga	Gaeilge, Gàidhlig na hAlban	Béarla	Gaeilge, Béarla	Béarla	Roinnt teangacha, Gaeilge san áireamh
Sprioctheanga	Gaeilge, Gàidhlig na hAlban	Gaeilge	Gaeilge, Béarla	Gaeilge	Roinnt teangacha, Gaeilge san áireamh
Réimse teanga	Ilfheidhmeach	Riarachán poiblí	Ilfheidhmeach	Riarachán poiblí	Ilfheidhmeach
Slat tomhais	Uathoibríoch	Uathoibríoch	Uathoibríoch & Suibhachtúil	Uathoibríoch	Anaithnid

Tábla 1: Meaisínaistriúchán na Gaeilge

De ghnáth, bíonn caighdeán an aschuir níos fearr ó chóras NMT i comparáid le córas SMT, ach ní gá go mbeadh sé sin fíor i gcomhthéacs na Gaeilge. Rinne Dowling *et al.* (2018) comparáid chuimsitheach idir SMT agus NMT sa chás sin agus forbraíodh dhá chóras inchomparáide. Chun comparáid réasúnta díreach a dhéanamh idir an dá chóras, úsáideadh an corpas comhthreomhar céanna chun na córais a chruthú: téacs ón Roinn Cultúir, Oidhreacht agus Gaeltachta; DCEP: Digital Corpus for European Parliament (Hajlaoui *et al.*, 2014); DGT-TM: Directorate General for Translation (Steinberger *et al.*, 2012). Baineadh úsáid as slat tomhais uathoibríoch agus fuarthas go raibh an caighdeán níos fearr ón gcóras SMT, is dócha go bhfuil níos mó téacs ag teastáil chun córas NMT den chaighdeán céanna a fhorbairt.

Conclúid

San alt seo pléadh cuid de na forbairtí i réimse an mheaisínáistriúcháin le blianta beaga anuas. Cé go bhfuil méadú as cuimse ar úsáid an mheaisínáistriúcháin ó d'fhorbair Google a gcóras féin, ní fhéadfaí a rá go bhfuil an teicneolaíocht ag dul i dtreo na foirfeachta sna mórtheangacha, gan trácht ar mhionteangacha ar nós na Gaeilge.

Is uirlis úsáideach atá sa mheaisínáistriúchán d'aistritheoirí na Gaeilge, fad is go bhfuil tuiscint acu ar conas é a úsáid, agus na laincisí a bhaineann leis. Tá córais atá dírithe ar réimse ar leith níos úsáidí (mar go bhfuil siad níos cruinne) ná córais ilfheidhmeacha – mar go bhfuil an corpas níos gaolmhaire le comhthéacs an téacs.

Leabharliosta

- Arčan, M. *et al.*, (2016) 'IRIS: English-Irish Machine Translation System.' *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia.
- Bojar, O. *et al.*, (2016) 'Findings of the 2016 Conference on Machine Translation.' *Proceedings of the First Conference on Machine Translation (WMT 2016), Volume 2: Shared Task Papers*. doi: 10.18653/v1/W18-64028.
- Church, K.W. & Hovy, E.H., (1993) 'Good applications for crummy machine translation.' *Machine Translation*. doi: 10.1007/BF00981759.
- Coughlin, D., (2003) 'Correlating Automated and Human Assessments of Machine Translation Quality.' *Proceedings of the MT Summit IX*.

- Dowling, M. *et al.*, (2015) 'Tapadóir: Developing a Statistical Machine Translation Engine and Associated Resources for Irish.' *The 4th LRL Workshop: 'Language Technologies in support of Less-Resourced Languages'*. Poznan.
- Dowling, M. *et al.*, (2016) 'English to Irish Machine Translation with Automatic Post-Editing.' <https://www.computing.dcu.ie/~ygraham/tapadoir-cltw.pdf>
- Dowling, M. *et al.*, (2018) 'SMT versus NMT: Preliminary comparisons for Irish.' *Proceedings of the AMTA 2018 Workshop on Technologies for MT of Low Resource Languages (LoResMT 2018)*. Association for Machine Translation in the Americas. 12–20. Le fáil ag: <http://aclweb.org/anthology/W18-2202>.
- Forcada, M.L. *et al.*, (2011) 'Apertium: A free/open-source platform for rule-based machine translation.' *Machine Translation*. doi: 10.1007/s10590-011-9090-0.
- Hajlaoui, N. *et al.*, (2014) 'DCEP: Digital Corpus of the European Parliament.' *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*.
- Kenny, D., (2012) 'The Ethics of Machine Translation.' *NZSTI Annual Conference*.
- Kim, Y., Huang, J. & Emery, S., (2016) 'Garbage in, garbage out: Data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection.' *Journal of Medical Internet Research*. *Journal of Medical Internet Research*, 18(2). doi: 10.2196/jmir.4738.
- Méchura, M.B., (2017) *An Ríomhaire Ilteangach*. Baile Átha Cliath: Cois Life.
- Moorkens, J. & Way, A., (2016) 'Comparing Translator Acceptability of TM and SMT Outputs.' *Baltic J. Modern Computing*.
- Offersgaard, L. *et al.*, (2008) 'Domain specific MT in use.' *EAMT 2008 12th annual conference of the European Association for Machine Translation September*. 22–3.
- Papineni, K. *et al.*, (2002) 'BLEU: a Method for Automatic Evaluation of Machine Translation.' *ACL*. doi: 10.3115/1073083.1073135.

- Scannell, K.P., (2006) 'Machine translation for closely related language pairs.' *Proceedings of the 5th SALTMIL Workshop on Minority Languages and the 5th International Conference on Language Resources and Evaluation (LREC-2006)*.
- Snover, M. *et al.*, (2006) 'A Study of Translation Edit Rate with Targeted Human Annotation.' *Proceedings of association for machine translation in the Americas*. doi: 10.1.1.129.4369.
- Steinberger, R. *et al.*, (2012) 'DGT-TM: A freely available Translation Memory in 22 languages.' *8th international conference on Language Resources and Evaluation (LREC)*. 454–9. doi: 10.1002/gcc.20795.
- Steinberger, R. *et al.*, (2014) 'An overview of the European Union's highly multilingual parallel corpora', *Language Resources and Evaluation*. doi: 10.1007/s10579-014-9277-0.
- Sun, Y., (2010) 'Mining the Correlation between Human and Automatic Evaluation at Sentence Level.' *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA).
- Turian, J.P., Shen, L. & Melamed, I.D., (2003) 'Evaluation of Machine Translation and its Evaluation.' *Proceedings of MT Summit IX*. doi: 10.1021/jp982896o.
- Wu, Y. *et al.*, (2016) 'Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.' <https://arxiv.org/abs/1609.08144>