

Leksis: Prawf Geirfa Ymaddasol Am Ieithoedd Cyfyngedig eu Hadnoddau

Alan Kersaudy



PRIFYSGOL
BANGOR
UNIVERSITY

Ysgol Iaith Diwylliant a Chelfyddydau
Coleg y Celfyddydau, y Dyniaethau a Busnes

Cyflwynwyd i fodloni'n rhannol â'r gofynion ar gyfer y
Radd Master of Science mewn Technoleg Iaith.

Arolygwraig Dr. G. Bovolenta

Medi, 2025

Datganiad o Wreiddioldeb

Mae'r gwaith a gyflwynwyd yn yr adroddiad hwn yn gyfan gwbl o astudiaethau'r myfyriwr unigol, ac eithr na hynny lle nodir. Lle cyflwynir dadleuon a'r tarddiad o'r gwaith yn gyfan neu'n rhannol o ffynonellau eraill, rhoddir cyfeiriad llawn i'r awdur gwreiddiol. Ni chyflwynwyd y gwaith hwn o'r blaen ar gyfer unrhyw radd, ac nid yw ar hyn o bryd o dan ystyriaeth gan unrhyw gorff sy'n dyfarnu graddiau arall.

Student:

Alan Kersaudy

Datganiad o Argaeledd

Adwaen yma bod unrhyw ran o'r adroddiad hwn ar gael i'w weld, ei ffotogopio neu ei chynnwys mewn astudiaethau yn y dyfodol, gan roi cyferiad llawn i'r tarddiad o unrhyw wybodaeth sydd wedi'i gynnwys yma.

Student:

Alan Kersaudy

Abstract

This dissertation introduces leksis, a lexical decision task vocabulary test. The scoring mechanism is derived from the Elo rating system, used in chess, which works on the same principles as the Rasch model more common in psychometrics. Several versions of the tests were implemented so far, a Breton one first, then others for Welsh, French and Ukrainian. The goal of the dissertation is to measure the adaptivity the tests. It is admitted that the purpose of the test is to bring the test taker to the point where the chances of recognising a real word are exactly at 50%, that is, perfectly uncertain. After a little more than two months, the results collected for the Breton test seem conclusive over several vocabulary level ranges. The ranges with the most collected data seem well calibrated, which seems to consolidate the idea that little calibration effort is needed to create a reliable test. The innovation making this feat possible is a method we dubbed “bean” or “modulo” clustering. This item rating initialisation technique effectively turn any fully adaptive Elo based test in a hybrid system between a proportion of correct answers based system and a logistic scale it is supposed to be in the first place, thus finding a trade-off between calibration of new items (scalability) and exploitation of known items (precision). The test also introduces the use of RNN (LSTM) in the making of pseudo-words, which appeared to be convincing, with a fifth of them being recognised as real words in average by the test takers. At the end of a session, the test creates a personalised analysis prompt that test takers can share with an LLM to generate an interactive language lessons based on the results. A tool whose pedagogical value is still to be assessed.

Cydnabyddion

Ni ddylid priodoli unrhyw ansawdd a geir yn y draethawd hwn i'w awdur, nad yw'n dwyn ond y cyfrifoldeb bychan o roi ynghyd syniadau, mewnwelediadau, a myfyrdodau eraill. Hyd yn oed cafodd yr ymdrech fechan hon ei chynnal yn uniongyrchol neu'n anuniongyrchol gan bobl a sefydliadau a wnaeth fy arhosiad ym Mangor mor ddymunol, buddiol a chynhyrchiol ag y gallai fod. Yn arbennig, rwy'n ddyledus yn uniongyrchol i Giulia Bovolenta, a'm cyfeiriodd gyntaf tuag at brofi geirfa, ac y bu ei mewnwelediadau niferus yn arwain pob cam drwy'r gwaith hwn. Maria Kolesnichenko yw goleudy'r synnwyr cyffredin caredig a'm cadwodd rhag boddi mewn môr o syniadau gwael drwy'r flwyddyn gyfan. Melanie Jouitteau, a blannodd gyntaf yn fy meddwl y syniad o wneud y gradd Meistr hon ym Mangor. Dysgu Philip Davies yn y cwrs TSI a'm helpodd i wella fy sgiliau Cymraeg, ynghyd â chyd-fyfyrwyr Cymraeg eu hiaith y gallaf ond bod yn ddiolchgar i'w galw'n ffrindiau: Leena Farhat, Tom Williams, Stephen Russel ac Owen Williams. Mae Preben Vanberg ymhlith llawer o bethau yn berson cyfrifol am y syniad o osod anhawster yr eitemau ar yr un raddfa logistaidd â'r rhai sy'n cymryd y prawf. Sarah Cooper am helpu cyfieithu'r gwaith hwn yn Gymraeg. Tîm cyfan Canolfan Bedwyr, yn eu plith Tegau, Cat, Alun, Steffano a llawer o rai eraill. Ond yn arbennig pennaeth ei Huned Dechnoleg Gruffydd Prys a'i deulu, na ellir gwneud cyfiawnder â geiriau i'w hymdrech i'm croesawu a chefnogi fy arhosiad ym Mangor. Dau athro arall, Dewi Bryn Jones a Bill Teahan, a'm helpodd, mewn ffordd wahanol iawn ond gyflenol, i feistroli llawer o'r sgiliau a wnaeth y gwaith hwn yn bosibl. Yn olaf, mae'r gefnogaeth ariannol a gefais gan Cymen yn gyfle a'm hysgogodd i gynhyrchu'r teclyn gorau y gallwn ei gynhyrchu'n faterol ar gyfer y Gymraeg. Yn y pen draw, rwy'n cyflwyno'r prawf hwn a'r asiant tiwtiora sy'n gysylltiedig ag ef i Cymen a'i chyfarwyddwyr Manon Cadwaladr ac Aled Jones yn ogystal â Myfyr Prys, am eu rôl yn galluogi ysgoloriaeth Cymen. Boed yn arwydd bach o ddiolchgarwch am yr ymddiriedaeth a'r rhyddid y maent wedi ei roi imi. Rwy'n fendigedig ac yn ddiolchgar am fod wedi croesi ffordd cymaint o bobl ddeallus a da ac mae'r gwaith hwn yn bennaf ffrwyth eu caredigrwydd cyfunol.

Cynnwysion

Rhestr o Ffigurau	vi
Rhestr o Dablau	vii
Rhestr o Dalfyriadau	viii
1 Rhagymadrodd	1
1.1 Strwythur y Traethawd	1
1.2 Nod, Amcanion a Chwestiwn Ymchwil	1
1.3 Cefndir a Chymhelliant	2
1.3.1 Terminoleg: DAmA ac EdTech	2
1.3.2 Ieithoedd Isadnodd mewn Technolegau Addysgol	2
1.3.3 Deallusrwydd Artiffisial ac Addysg	4
1.3.4 Addasrwydd a Modelau Gwybodaeth	5
1.3.5 Casgliad	7
2 Adolygiad Llenyddiaeth	8
2.1 Lluniadau'r Hyfedredd a Ble i'w Canfod	8
2.1.1 Y Dullau Cyfannol o Brofi (CEFR)	9
2.1.2 Natur gymhleth y syniad o Hyfedredd	10
2.1.3 Topograffi y Profion Geirfa	13
2.1.4 Perthnasedd a Chyfngiadau Profion Geirfa	15
2.1.5 Casgliad	16
2.2 Olrhain Gwybodaeth	17
2.2.1 Gallu Damcaniaethol Prawf Unddimensiwn Diswn	18
2.2.2 Y System Graddio Elo	19
2.3 Casgliad	21
3 Methodoleg	23
3.1 Ffynonellu'r Allweddiau	23
3.2 Cynhyrchu'r Llithiau	24
3.2.1 Hyfforddi'r Model	24
3.2.2 Cynhyrchu Geiriau Ffug	25

3.3	Rhoi Gradd Cychwynol i'r Eitemau	27
3.3.1	Sut mae'r Graddau Cychwynol yn Effeithio ar Addasolrwydd	27
3.3.2	Bagu Modwlo	27
3.3.3	Gradd Cychwynol yr Llithau	28
3.4	Byrhau'r Rhestr o Eitemau	29
3.5	Y Sesiynau	30
3.5.1	Diweddarau Sgôr y Defnyddwyr a Hyd y Sesiwn	30
3.5.2	Dewis yr Eitemau mewn Sesiwn	31
3.6	Diweddarau Sgôr yr Eitemau	31
3.7	Ychwanegu Ieithoedd Newydd	32
3.8	Atborth	33
3.9	Dilysu'r Prawf	33
3.9.1	Dilysrwydd Lluniad a Dewisiadau Dylunio	33
3.9.2	Eglurdeb ar Ddehongli Canlyniadau	36
3.9.3	Protocol Dilysu	36
4	Canlyniadau	38
4.1	Descriptive Statistics	38
4.2	Mesur Addasu	40
5	Trafodiaeth	42
5.1	Y Prawf ar Arfer	42
5.1.1	Oed a Chysylltiad â Risg	42
5.1.2	Gwerth y Prawf fel Offeryn Dysgu	43
5.2	Cyfyngiadau a Dehongliad y Sgôr	44
5.3	Ymchwil Dyfodol	44
5.4	Casgliad	45
A	Prompt Dadansoddi	47
A.1	Patrymlun	47
A.2	Enghraifft	48
	Llyfryddiaeth	51

Rhestr o Ffigurau

2.1	Fformiwla Rasch	19
2.2	System graddio Elo	19
3.1	Gwahaniaeth mewn dosbarthiad y characteriau (geiriau ffug / geiriau go iawn) .	26
3.2	Gwahaniaeth rhwng y cyfiri o eiriau ffug dros eiriau go iawn ar raddfa logarifmig ar gyfer hyd penodol.	26
3.3	Yr un gwahaniaeth ag yn 3.2 wedi'i rannu trwy'r cyfanrif o eitemau i ddod â'r gwahaniaethau yn y cyd-destun o sesiwn brofion.	26
3.4	Dosbarthiad yr eitemau ar ôl ddechrau'r proses calibri	28
3.5	Sgrinlun o rwyngwyneb y brawf yng nghabol sesiwn brawf Lydaweg.	30
3.6	Sgrinlun o'r sgrin olaf ar ôl cwblhau sesiwn brawf, gyda'r opsiynau i rannu canlyniadau neu gael adborth gan LLM.	34
4.1	Dosbarthiad sgôrion ar draws sawl prawf a'u cyfraddau ffug positif cysylltiedig	39
A.1	Dechrau gwers bersonol 1/2	49
A.2	Dechrau gwers bersonol 2/2	50

Rhestr o Dablau

3.1	Categoriâu o eiriau a dynnwyd o Devri (wedi'u hidlo) a Meurgorf	23
4.1	Cyfartaledd y gair go iawn olaf a adnabuwyd mewn ystodau penodol o sgôrion terfynol y prawf Llydaweg, gyda gwerthoedd-p o dest binomial	41

Rhestr o Dalfyriadau

CAC	Cyfran o Atebion Cywir (neu PC am Proportion of Correct Answers)
CAI	Caffael Ail Iaith
CDG	Cynnyrch Domestig Gros
CEFR	Fframwaith Cyfeirio Cyffredin Ewropeaidd ar gyfer ieithoedd (Common European Framework of Reference for languages)
DA	Deallusrwydd Artiffisial
DAmA	Deallusrwydd Artiffisial mewn Addysg (AIED, AIED)
DD	Dysgu Dwfn
EdTech	Technolegau Addysg
GD	Gwyddor Dysgu
GOFAI	Good Old-Fashioned AI (Ddeallusrwydd Artiffisial Da Hen-Ffasiynol)
HAdd	Holiad Amlddewis
IAC	Ieithoedd Adnoddau Cyfyngedig
IAL1	Iaith Adnoddau Llawn
L1	Iaith Gyntaf neu Brifiaith
L2	Ail Iaith (ieithoedd)
LDT	Lexical Decision Task – Tasg Penderfynu Geirfaol
LLM	Large Language Model
LSTM	Long Short-Term Memory
OG	Olrhain Gwybodaeth
PAC	Profi Addasol Cyfrifiadurol (neu CAT am Computerized Adaptive Test)

PIN	Prosesu Iaith Naturiol
RNN	Recurrent Neural Network
SDT	Theori Canfod Signalau
TYE	Theori Ymateb Eitemau (neu IRT am Item Response Theory)
VLT	Vocabulary Level Test – Prawf Lefelau Geirfa
WEIRD	Gorllewinol, Addysgedig, Diwydiannol, Cyfoethog a Democrataidd (o'r Saesneg)

Pennod 1

Rhagymadrodd

Mae'r bennod gyntaf hon yn cyflwyno'r cyd-destun, a phwrrpas y traethawd hir hwn. Yn benodol, daw'r drydedd adran â'r rôl y sydd gan dechnolegau addysgol i'w chwarae i'r amlwg, naill ai wrth gefnogi neu beryglu ymhellach Ieithoedd Adnoddau Cyfyngedig (IAC), yn dibynnu ar ai yw'r dechnoleg yn cael ei chynllunio i ddysgu ieithoedd sydd eisoes yn bygwth ieithoedd eraill. Gellir darllen yr adran hon fel cyflwyniad cyffredinol i faes technolegau addysgol i'r rhai hynny sy'n pryderu ynghylch tynged ieithoedd adnoddau cyfyngedig, neu fel cyflwyniad i bryderon ieithoedd adnoddau cyfyngedig i'r rhai hynny sy'n gweithio ym maes y technolegau addysgol.

1.1 Strwythur y Traethawd

Mae'r traethawd hwn yn cyflwyno Leksis, sydd yn brawf adnabod geirfa newydd wedi ei deilwra ar gyfer ieithoedd adnoddau cyfyngedig. Mae'r bennod gyntaf yn egluro'r rhesymeg tu ôl i brawf o'r fath. Daw'r ail bennod â rhannau o'r llenyddiaeth sydd ar gael o wahanol feysydd at ei gilydd, gan amrywio o ieithyddiaeth gymhwysol i ddamcaniaeth gwybodaeth (information theory), er mwyn gosod y sylfaen ar gyfer profion geirfa graddadwy sydd yn addasedig i gyfyngiadau a chyd-destun yr ieithoedd adnoddau cyfyngedig. Mae'r drydedd bennod yn cyflwyno cynllun prawf cychwynnol ar gyfer y Llydaweg. Y bedwaredd bennod y sydd yn dadansoddi canlyniadau'r brawf i asesu perthnasedd y dewisiadau dylunio. Yn olaf, mae'r bumod bennod yn asesu gwerth a chyfyngiadau'r brawf, yn ogystal â chyflwyno barn wybodus ar yr anghenion penodol i ieithoedd adnoddau cyfyngedig mewn perthynas â thechnolegau addysgol ac ieithyddol.

1.2 Nod, Amcanion a Chwestiwn Ymchwil

Mae ieithoedd cyfyngedig eu hadnoddau yn wynebu heriau arbennig mewn byd lle mae gwyddor data wedi troi maint yn fam pob ansawdd. Nod sylfaenol y gwaith presennol yw optimeiddio addysgu ieithoedd hyn. Problem hanfodol unrhyw optimeiddio yw'r metrig y

mae rhywun yn anelu at ei optimeiddio, ac mae hyn yn arwain at ddatblygu profion iaith cyflym, lleiafol a gyflwynir yma. Yn arbennig, yr amcan yw canfod ffyrdd o wneud iawn am broblem prinder adnoddau trwy ddatblygu dulliau a thechnegau wedi'u cynllunio i weithio yn y cyd-destun prinder hwn yn gyntaf, yn hytrach na chludo i ieithoedd prin eu hadnoddau ddulliau a thechnegau sy'n rhy drwm ar adnoddau. Am y rhesymau hyn, rydym yn cynnig y cwestiwn ymchwil canlynol.

A ellir creu prawf geirfa cyflym, ymaddasol ar gyfer ieithoedd prin eu hadnoddau?

Mae ymaddasoldeb yn crynhoi dibynadwyedd a manwl gywirdeb, dau agwedd allweddol ar ddilysrwydd seicofetrig. Felly gellir defnyddio dilysu ymaddasoldeb fel procsi i ddilysu'r potensial i brawf olrhain cynnydd dros amser, heb astudiaeth ar raddfa fawr. Astudiaethau a fyddai'n gofyn am ddilyn grwpiau cyfan o ddysgwyr dros sawl mis i gasglu'r canlyniadau angenrheidiol. Trwy astudio'r llenyddiaeth sydd ar gael yn drwyadl ac asesu ymaddasoldeb prawf gwirioneddol o'i ganlyniadau cynnar, y bwriad yw cefnogi dadl gadarn o blaid neu yn erbyn y syniad hwn erbyn diwedd y gwaith hwn.

1.3 Cefndir a Chymhelliant

1.3.1 Terminoleg: DAmA ac EdTech

Mae ymchwil academiaidd fodern ar dechnolegau addysgol yn bennaf yn disgyn dan yr ymbarél “Deallusrwydd Artiffisial mewn Addysg” (DAmA neu AIED am “AI in Education”). Mae'r derminoleg hon yn dominyddu'r maes oherwydd “Cymdeithas y DAmA Ryngwladol” (International AIED Society), a sefydlwyd ym 1993, ac effaith strwythurol rhifynnau ei chyfnodolyn a'i chynadleddau. Yma, defnyddir DAmA ar brydiau bron yn gyfnewidiol ag EdTech, am “Technolegau Addysgol”, sef term mwy o'r maes marchnata, term sy'n perthyn yn agosach at neolegolion eraill fel “FinTech”, “BioTech” ac ati. Gellir ystyried cwmnïau addysgol fel Duolingo neu Rocket Language fel cwmnïau EdTech yn ôl y diwydiant, ond fel DAmA pan siaradir gan ymchwilwyr. Mewn fformiwleiddiad arall, EdTech yw'r DAmA gyda model busnes.

1.3.2 Ieithoedd Isadnodd mewn Technolegau Addysgol

Mae cwestiwn ieithoedd adnoddau cyfyngedig yn y DAmA yn gysylltiedig yn agos â'u sefyllfa gyffredinol ym maes prosesu iath naturiol (PIN). Disgrifir y sefyllfa yn orau yn Magueresse, Carles and Heetderks (2020), wrth i ddulliau ystadegol, cysylltiadol, ddod i

ddominyddu mewn PIN, mae cwestiwn prinder data yn dod yn brif ffactor cyfyngol wrth gymhwyso atebion PIN modern i ieithoedd adnoddau cyfyngedig. Mae'r broblem hon hefyd yn cael ei chymhlethu gan duedd WEIRD (am Wealthy, Educated, Industrialized and Democratic) cyffredinol mewn gwyddor wybyddol (Henrich, Heine and Norenzayan 2010), lle mae ieithoedd o ddiwylliannau sy'n orllewinol, addysgedig, diwydiannol, cyfoethog a democrataidd yn tueddu i gael eu ffafrio ym mhob maes o'r gwyddor gwybyddol. Fodd bynnag, os mai'r ieithoedd adnoddau cyfyngedig sy'n mabwysiadu'r technolegau hyn y lleiaf, yn eironig yr ieithoedd hyn y sydd â'r mwyaf i'w golli o beidio â'u mabwysiadu. Gall peidio â mabwysiadu'r technolegau hyn achosi colli gwelededd, parch a dymunoldeb, sy'n ei dro yn arwain at lai o fabwysiadu a defnydd, gan arwain at gylch cythreulig lle mae llai o adnoddau hyfforddi ar gael i addasu'r technolegau hyn i ieithoedd cyfyngedig eu hadnoddau. Disgrifiwyd y ffenomen hon fel marweidd-dra (*stagnation* yn Saesneg), neu farwolaeth digidol iaith, sef llofnod ar-lein ieithoedd sydd wedi darfod yn gymdeithasol (Kornai 2013).

Ni ellir tanbrisiu'r rôl y gallai technolegau addysgol ei chwarae wrth dorri'r cylch cythreulig hwn, o leiaf ar gyfer rhai o'r ieithoedd dan sylw. Ar y naill law, gall helpu i addasu technolegau addysgol presennol i ieithoedd adnoddau cyfyngedig helpu i gynnal eu perthnasedd fel cyfrwng dysgu. Hynna i'r rieni sy'n dymuno'r safonau addysgol gorau i'w plant yn ogystal a chynnig dewis arall i bobl sy'n ceisio cyflawniad deallusol yn hytrach na gadael eu prifiaith yn syth i allu parhau i ddysgu pethau newydd. Deallir y gall technolegau PIN fel cyfieithu awtomatig helpu i drosglwyddo technolegau i gymunedau ieithyddol na all ddatblygu eu hofferynnau addysgol eu hunain fel arall (Haddow et al. 2022). Mae astudiaeth gan Horbach et al. (2024) yn cefnogi'r syniad y gellir cyflawni cydraddoldeb addysgol trwy systemau sgorio traws-ieithog, yn y cyd-destun lle defnyddir cwestiynau agored i asesu sgiliau, a lle gall cefndiroedd ieithyddol gwahanol effeithio ar ruglder atebion myfyrwyr beth bynnag yw eu dealltwriaeth o'r cysyniad a asesir. Ar y llaw arall, o ran technolegau addysgol sy'n gyfeirio at iaith yn benodol, mae'r maes bron i gyd yn dominyddu ymchwil i ddysgu Saesneg, a hyd yn oed yn dod i gystadlu ag ieithoedd sydd eisoes mewn perygl. Mae papur gan Henkel et al. (2025) yn symptomatig o'r peryglau hynny. Yn yr astudiaeth hon, defnyddir technolegau adnabod lleferydd Saesneg mewn system DAmA i wella llythrennedd mewn ysgolion Ghana, gwlad sy'n gartref i fwy na 70 o ieithoedd brodorol (*Ghana Languages, Literacy, Maps, Endangered Languages, Population, Official Use (GH)* 2025).

Hyd y gwyddom, ymddengys mai ychydig o ymdrech a wnaed yn y llenyddiaeth academiaidd i gefnogi datblygu technolegau addysgol wedi'u teilwra'n benodol ar gyfer anghenion ieithoedd adnoddau cyfyngedig a'u cymunedau siaradwyr, er gwaethaf yr holl gynnydd a wnaed yn y blynyddoedd diwethaf i ddatblygu'r ieithoedd hyn mewn PIN. Gallai'r diffyg tystiolaeth hwn fod wedi'i achosi gan rwystr iaith, ond nid yw hyn ond yn atgyfnerthu'r syniad y dylid, os nad oes rhaid, gwneud mwy i gefnogi presenoldeb yr ieithoedd adnoddau cyfyngedig yn y DAmA.

1.3.3 Deallusrwydd Artiffisial ac Addysg

Fel y dangoswyd gan Doroudi (2023), bu rhwng ddeallusrwydd artiffisial (DA) ac ymchwil mewn addysg ddeialetig 70 mlynedd o hyd a fu o fudd i'r ddau faes hyn. Os tynnodd gwaith cynnar ar DA o seicoleg ddatblygiadol yn wreiddiol a hyd yn oed ddatblygu offer addysgol fel rhan o'u hymdrech i efelychu deallusrwydd dynol gyda pheiriannau, maes addysg sydd bellach yn elwa o'r posibilïadau a ddatgloir gan dechnolegau DA modern.

Archwiliodd ymchwil gynnar mewn deallusrwydd artiffisial ddau ddull gwahanol i geisio efelychu prosesau gwybyddol. Gelwir y cyntaf yn gyffredin fel Good Old-Fashioned AI (GOFAI), roedd wedi'i ganoli o amgylch dull symbolig a ddeilliodd o waith semenaidd Allen Newell, Herbert A. Simon a Cliff Shaw ar y Logic Theorist (Newell and Simon 1956). Ceisiodd y dull hwn ddeall sut mae arbenigwyr yn datrys problemau gan ddefnyddio systemau sy'n seiliedig ar reolau a haniaethu symbolig. Roedd yr ail ddull, cysylltiadol, wedi'i ganoli o amgylch rhwydweithiau niwral ac yn canolbwyntio ar y prosesau caffael sgiliau gwybyddol dros berfformiad priodol. Yn y byd Cymraeg, mae Cysill (Hicks 2004) yn engraifft o system DA seiliedig ar reolau. Datblygwyd GOFAI gan bobl fel Marvin Minsky, Seymour Papert a llawer o rai eraill (Doroudi 2023). Yn nodedig, daeth Seymour Papert i'r byd DA ar ôl astudio datblygiad gwybyddol plant yn labordy Jean Piaget yn Geneva. Daeth â dylanwad sylweddol o adeiladaeth (constructionism) Piaget i'r paradigm cysylltiadol mewn DA, sy'n ddamcaniaeth a esyd bod dysgwyr yn adeiladu eu sgiliau newydd a'u dealltwriaeth ar ben y wybodaeth a'r sgiliau a gafwyd eisoes.

Arweiniodd y ddau ddull at ymdrechion i greu systemau addysgol rhyngweithiol yn gynnar. Mae enghreifftiau o raglenni meddalwedd addysgol cynnar sy'n seiliedig ar GOFAI yn cynnwys system GUIDON, a oedd yn dibynnu ar beiriant Mycin, system ddiagnoseg haintiau, i ddysgu myfyrwyr i ddiagnosio patholegau (William J. 1983). Cefnogodd y gangen

gysylltiadol ddatblygu “micro-byddydd” addysgol, megis ieithoedd rhaglennu addysgol, lle gallai plant ddysgu sgiliau datrys problemau anniffiniedig. Mae enghreifftiau o’r dull hwn yn cynnwys iaith raglennu Logo, a ddyluniwyd i ddysgu am leoli cymharol a geometreg trwy ddylunio rhaglenni i arwain crwbanod robot (darllunio). Dilynodd llawer o systemau o’r fath, megis iaith raglennu Scratch a chitiau Lego Mindstorms. Ond arweiniodd yr arbenigedd angenrheidiol mewn DA at ymchwil ddiweddarach yn canolbwyntio’n llwyr ar berfformiad systemau cyfrifiadurol, yn enwedig wrth i ddyfodiad yr ôl-ledaeniad (back-propagation) esgor dysgu dwfn (DD) a llwyddo i sefydlu uchafiaeth y paradigm cysylltiadol mewn DA.

Ar y pwynt hwn, symudodd y ffocws yn bendant o ddefnyddio seicoleg ddatblygiadol i gefnogi DA, i integreiddio atebion technegol DA mewn offer addysgol. Mae meta-ddadansoddiad gan Schmid et al. (2023) bellach yn cefnogi manteision dulliau addysgol adeiladol fel Dysgu Cyfun (Blended Learning) a’r Ystafell Ddosbarth Wedi’i Throi (Flipped Classroom), y sy’n rhoi mwy o rôl hyfforddi i’r athrawon, gyda chyfrifoldeb cyfarwyddo yn cael ei drosglwyddo i systemau rhyngweithiol ar-lein, a ddefnyddir tu allan i’r ddosbarth.

Yn yr adran hon, gwelwyd sut y llifodd syniadau adeiladol Piaget ar addysg yn y ddull gysylltiadol at y DA trwy waith Seymour Papert. Yna, pan gymerodd y dull cysylltiadol drosodd gyda dyfodiad y DD, daeth y DA yn ôl i’r addysg ar ffurf platfformau dysgu addasol i gefnogi datblygiad arferion adeiladol mewn ysgolion. Mae dysgu am yr hanes cyfun hwn yn mynd y tu hwnt i ymholiad am straeon hanesyddol yn unig, mae’n rhoi inni’r cwmmpas a’r fframwaith epistemolegol i bennu nodau a dulliau technolegau addysgol, sy’n gam angenrheidiol i sicrhau y gallai systemau addysgol newydd o’r fath gyflawni llwyddiant byd-go-iawn rywbyrd. Hynny yw, nid fel system ynysig sy’n esblygu mewn gwactod, ond fel offer yng ngwasanaeth amgylchedd dysgu cyfannol.

1.3.4 Addasrwydd a Modelau Gwybodaeth

Addewid Addasrwydd

Y gwahaniaeth allweddol rhwng gwersllyfrau clasurol neu addysg sy’n seiliedig ar ddarlithiau a’r rhan fwyaf o’r technolegau dysgu diweddar yw addewid addasrwydd. Mae hyn yn golygu bod y system yn addasu ei hymddygiad yn seiliedig ar berfformiad y dysgwyr, yn ddelfrydol gyda’r nod o fwyhau eu derbyniad dysgu. Yn y rhan fwyaf o systemau modern (ond nid pob un) gwneir y optimeiddiad hwn gan system argymell, y mae ei ffurfiau mwyaf soffistigedig

yn datrys enghraifftiau o broblem y bandit aml-fraich¹. Problem y bandit aml-fraich sef fformiwleiddiad mathemategol o sefyllfa lle cynigir gweithredoedd gwahanol, yn ein hachos ni, argymell deunyddiau dysgu gwahanol gyda gwerthoedd pedagogig ansicr, a rhaid i asiant benderfynu pa weithredoedd fydd yn mwyhau enill ar wobr benodol, yma, twf y myfyrwyr mewn gwybodaeth. Gellir ei datrys gan un o sawl algorithm gwahanol (Chen et al. 2017). Rhaid i'r systemau hyn dewis rhwng ecsbloetio gweithredoedd gyda gwobrau hysbys, ond cyfyngedig ac archwilio gweithredoedd gyda gwobrau anhysbys.

Mae'r paradigm hwn yn galluogi dylunwyr systemau i ryddhau eu hunain o'r pen tost a achosir gan yr angen i ddyfarnu'r cwestiwn sy'n ymwneud â dewis y deunydd dysgu. Yn dibynnu ar yr algorithm a ddewiswyd, addewid dysgu addasol yw galluogi adeiladu proffil wedi'i deilwra am sgiliau'r dysgwyr, o bosibl hefyd yn cynnwys disgrifiad o'u gallu neu rhythm dysgu, a chael y system i adeiladu cwricwlwm wedi'i optimeiddio i gyrraedd y nod pedagogig penodedig.

Rhaid nodi bod technolegau addysg seiliedig ar reolau yn dal cael eu gweithredu'n eang, lle mae'r cwricwlwm yn cael ei ddylunio o flaen llaw ar sail model pedagogaidd, ac felly yn cymryd rôl y systemau argymell a gyflwynwyd uchod. Gall y systemau hynny fod yn berthnasol pan mai'r nod yw dysgu setiau penodol, diffiniedig o sgiliau, fel rhaglenni ysgolion cynradd ac uwchradd. Mae Pelánek (2025) yn crybwyll y platfform *Umíme* yng Ngweriniaeth Tsiec, sy'n ymddangos fel ei fod wedi'i fabwysiadu'n helaeth gan ysgolion yr wlad ac yn dibynnu ar bensaernïaeth o'r fath. Efallai bydd gan systemau addysgol eraill priodweddau rhyngweithiol yn unig, heb systemau argymell, fel yr ieithoedd rhaglennu addysgol a grybwyllwyd uchod. Fodd bynnag, nid y rheiny yw ffocws y gwaith presennol.

Model Gwybodaeth a Nod Offerynnol

Lle mae systemau argymell yn gallu gwneud yr addewid i optimeiddio unrhyw fetrig penodedig, o amser gwyllo fideo YouTube i weithgynhyrchu clipiau papur (Bostrom 2003), nid yw systemau DA yn dwyn y cyfrifoldeb i ddiffinio'r metrigau cyfryngol hyn, yr hyn a elwir yn nod offerynnol. Mae'r cwestiwn hwn wrth wraidd pob ystyriaeth aliniad, ac nid yw'r systemau addysgol yn ddieithr i'r broblem hon. Mewn systemau addysgol, mae'r procsi hwn yn seiliedig ar fodelau gwybodaeth, a elwir hefyd yn fodelau myfyriwr, sy'n ddata seicometrig y gellir deillio model dysgu ohono (sef y datblygiad o'r wybodaeth honno ar y amser) ac a

¹Multi-armed bandit, cf. https://en.wikipedia.org/wiki/Multi-armed_bandit

all yn ei dro gael ei ddefnyddio i ddiffinio gwerth pedagogig deunydd dysgu, y metrig hwn yw'r wobr y byddai algorithmau yn cael eu cyhuddo o'i optimeiddio. Mae diffinio'r model gwybodaeth hwn a natur y lluniad seicometrig y mae'n ei gasglu yn hanfodol i lwyddiant system dysgu addasol. A'r gwaith deffinio hwn sy'n gyfrifoldeb y maes y bwriedir i'r system ei ddysgu, yn ogystal modelau seicolegol, nid y technoleg yn uniongyrchol.

1.3.5 Casgliad

Trwy'r adran hon, dadansoddwyd hanes technolegau addysgol ers y chwyldro gwybyddol yn y 1950au. Gwelwyd y potensial amhrisiadwy'r DAmA sy'n dal i ddatblygu, a'i addewid o addasrwydd, ynghyd â'r risgiau a'r cyfleoedd y mae'n eu dod i ieithoedd adnoddau cyfyngedig. Nodwyd bwlch yn y llenyddiaeth academiaidd ar ddysgu ieithoedd adnoddau cyfyngedig yn y DAmA. Os gall cyfieithu systemau DAmA i ieithoedd adnoddau cyfyngedig weithio cyn belled â bod y pwnc y bwriedir iddo ei ddysgu ddim yr iaith ei hun, pan fwriedir i'r system dysgu ieithoedd, mae'n ymddangos bod yr holl ymchwyl yn y maes wedi'i ddominyddu gan Saesneg, iaith sy'n eithriad o ran argaeledd adnoddau o'i gymharu â mwyafrif y 7000 o ieithoedd eraill a siaredir ledled y byd. Yn y cyd-destun hwn, mae'n ymddangos yn angenrheidiol aillfeddwl sut y gellir cyflawni addasrwydd pan nad oes gan y rhan fwyaf o ieithoedd y byd hyd yn oed ramadeg ddisgrifiadol wyddonol, heb sôn am y dwysau o oriau o recordiadau wedi'u hanodi sy'n angenrheidiol i hyfforddi systemau adnabod lleferydd.

Pennod 2

Adolygiad Llenyddiaeth

Mae'r adolygiad llenyddiaeth hwn wedi'i rannu'n ddwy brif adran. Mae'r adran gyntaf yn canolbwyntion ar ddadansoddi'r lluniadau a archwiliwyd wrth asesu hyfedredd iaith, ac ymhlith y rheini, pa rai a allai weithio mewn system ddysgu addasol, tra bo'r ail adran yn ymdrîn â'r ffyrdd ystadegol i sgorio rhyw luniad penodol. Y brif-egwyddor trwy'r bennod hon fydd symlrwydd y datrysiadau a gynigir, gan mai bob amser yw'n haws trwsio diffygion systemau syml na rhai systemau cymhleth.

2.1 Lluniadau'r Hyfedredd a Ble i'w Canfod

Dangosodd y rhagymadrodd fod diffiniad yr amcanion offerynnol y mae'n rhaid i system argymell (recommender system) eu hoptimeiddio yn perthyn i'r faes arbenigedd sy'n ymwneud â phrif nod y system, yn hytrach nag i'r dechnoleg ei hun. Mae profi iaith yn draddodiadol wedi bod yn fater ymchwil Caffael Ail Iaith (CAI), y gellir ei hystyried fel is-faes gwyddor dysgu (GD), ond mae'r maes hwn yn derbyn mewnbwn gan – ac yn perthyn yn agos i – seicoieithyddiaeth, ieithyddiaeth gymhwysol, ac fel y gwelwn, niwrowyddoniaeth. Dibynnir ar y rhain am ddealltwriaeth gyffredinol o'r prosesau sy'n gysylltiedig â defnydd a chaffael iaith. Heb honni bod yn adolygiad cynhwysfawr, bydd yr adran hon yn ceisio darparu trosolwg o hyfedredd iaith a'r ffyrdd i'w fesur.

Sut mesuro hyfedredd iaith? Mae'r cwestiwn hwn wedi cael ei astudio'n eang o fewn fframweithiau damcaniaethol amrywiol ac ar gyfer sawl pwrpas ymarferol. Y fwyaf pwysig efallai, ydy bod myndeiaid i bethau fel dinasyddiaeth, addysg neu swyddi newydd yn dibynnu ar feistrioliaeth iaith. Mynediadau sy'n gyfleoedd llunio bywyd, sydd wedi gwneud ei dilysiad yn fater symudedd cymdeithasol. Yn yr adran hon, cyflwynir y ffordd cyffredinolaf ac sydd yn cael ei defnyddio mwyaf i asesu sgiliau iaith, cyn symud tuag at atebion amgen a fyddai'n cyd-fynd ag anghenion system ddysgu addasol raddadwy. Yn olaf, asesir yr atebion amgen

hyn yn feirniadol. Mae'r ail adran yn canolbwyntio ar ddod o hyd i ffyrdd i fynd i'r afael â diffygion yr atebion amgen hyn.

2.1.1 Y Dullau Cyfannol o Brofi (CEFR)

Gellir asesu nodweddion cudd cymhleth fel hyfedredd iaith gan ddau baradigm profi, y cyntaf yn cael ei ddisgrifio fel uchafsymiol, cynhwysfawr neu gyfannol, a'r ail fel lleiafsymiol, ar sail-procsi neu leihaol. Mae profion iaith masnachol a sefydliadol megis yr IELTS a Chymwysterau Saesneg Caergrawnt ar gyfer Saesneg neu'r DELF a DALF ar gyfer Ffrangeg, yn dilyn dull uchafsymiol a ddiffinnir gan y Fframwaith Cyfeirio Cyffredin Ewropeaidd ar gyfer ieithoedd (CEFR) (Europe 2020). Nid yn unig y mae'r fframwaith hwn yn diffinio'r chwech gradd alffarifol enwog bellach o feistrolaeth iaith, ond hefyd y pedwar cyd-destun defnydd y dylid eu mesur ynddynt. Y pedwar hyn sydd yn gyfuniadau o ddau ddull defnydd, llafar ac ysgrifenedig, ar gyfer dau fath o weithgareddau, derbyn a chynhyrchu. Mae'n mesur y wybodaeth ieithyddol (geirfa, gramadeg a'u cydrannau) yn syth trwy'r pedwar sgîl iaith y mae defnyddwyr iaith yn ymgysylltu â hwy yn ymarferol: gwranddo, siarad, darllen ac ysgrifennu. Ystyrir y fframwaith hwn yn safonol y tu hwnt i ffiniau Ewrop, ond er gwaethaf ei gryfderau, efallai na fydd yn addas ar gyfer anghenion profi pob iaith.

Y brif feirniadaeth y gellid ei chodi yn erbyn y paradigm profi hwn yw'r ffaith mai dim ond deg iaith Ewropeaidd sy'n gallu ymfalchïo bod ganddynt brofion sy'n cydymffurfio â CEFR ac sy'n cynnwys y chwech lefel hyfredeedd y mae'n eu diffinio (*Common European Framework of Reference for Languages 2025*; *Cadre européen commun de référence pour les langues 2025*). Ar ôl pum mlynedd ar hugain o fodolaeth, nid yw hyd yn oed ieithoedd cenedlaethol economïau arweiniol yr UE fel Iseldireg neu Tsieceg yn perthyn i'r rhestr hon. Mae hwn yn ddiffyg sylfaenol ar gyfer paradigm a ddyluniwyd yn benodol i beidio â ffafrio prif ieithoedd yr Undeb. Mae'r rhesymau am hyn yn amlwg, dim ond yr ieithoedd mwyaf "marchnadadwy" sy'n gallu datblygu ecosystem addysgol digon cryf i wneud y profion hyn yn economaidd hyfyw. Weithiau, fe gall ewyllys wleidyddol pontio'r bwlch fel ar gyfer ieithoedd rhanbarthol yn Sbaen (mae Galisieg a Chatalaneg ymysg y deg iaith a grybwyllwyd uchod, pan mai dim ond prawf ar gyfer lefelau A1–2 sy'n eisiau ar Fasgeg), ond adeiladir yr ewyllys hon ar sefydliadau ac arbenigedd cryf sydd gan ddim ond llond llaw o ieithoedd ar gael iddynt yn Ewrop, heb sôn am weddill y byd. Er gwaethaf ei sylfaen ddamcaniaethol, mae prinder yr adnoddau (amser, arian, arbenigedd a diddordeb gwleidyddol) yn gwneud y paradigm cynhwysfawr profi ieithoedd yn anymarferol ar gyfer y rhan fwyaf ohonynt, ac y rhain sydd

unwaith eto'n cael eu gadael ar ôl. Unwaith eto, yr ieithoedd sydd â'r mwyaf i'w ennill o'r offer hyn, a'r mwyaf i'w golli trwy beidio â'u defnyddio, sy'n wynebu'r anawsterau mwyaf i gael mynediad atynt. Eto, yn achos system ddysgu addasol, sef y prif bwrpas ar gyfer y traethawd hwn, byddai prawf cynhwysfawr rhy hir ac yn diangen, gan y byddai'r profi yn cymryd gormod o amser o'r profiad dysgu, oni bai fod y profi yn rhan o'r addysgeg ei hun.

Rhaid inni edrych felly ar ffyrdd mwy effeithlon o fesur hyfedredd. Ond cyn hyn, mae angen inni ddatblygu dealltwriaeth ddyfnach o beth mae caffael iaith yn ei olygu. Sut y mae'r wybodaeth ddamcaniaethol haniaethol sy'n bresennol yn y geiriaduron a'r gramadegau nas darllenwyd gan neb yn cysylltu â'r dau neu bedwar sgîl ymarferol sy'n hysbys ledled y byd. Beth yw cymhwysedd a pherfformiad pan sonnir am hyfedredd iaith?

2.1.2 Natur gymhleth y syniad o Hyfedredd

Mae'r rhan fwyaf o ddamcaniaethau mewn ieithyddiaeth, yn enwedig strwythuriaeth de Saussure a chynrychioliaeth Chomsky, yn seiliedig ar ddull dadansoddol. Gan gymryd iaith ar wahân yn gyntaf o brosesau gwybyddol eraill, yna gwahanu ei chydrannau cysyniadol, geirfa oddi wrth ramadeg, cymhwysedd oddi wrth berfformiad (Chomsky 1965) a chadw ailadrodd y broses gyda'u cydrannau ac is-gydrannau. Ac yna, astudio'r ffyrdd i'w cyfuno gyda'i gilydd. Mewn ffordd, mae paradigm CEFR yn dilyn yr un tuedd epidemiolegol "dadansoddol", drwy rannu sgiliau cynhyrchu a chanfod, yn ogystal a defnydd llafar ac ysgrifenedig. Mae prif fantais y dulliau dadansoddol hyn yn amlwg, trwy wahanu agweddau a chategorïau, gall rhywun gyflawni dealltwriaeth gynhwysfawr o gydrannau a rheolau systemau cymhleth megis ieithoedd. Ond er gwaethaf ei gryfder, mae'r dull dadansoddol hwn yn dod â golwg ragfarnllyd o beth yw iaith, gan ei fod yn dod â llun statig ac ynysig i'r systemau y mae'n eu hastudio. Fodd bynnag, nid yw ieithoedd, neu ran hynny gwybodaeth ieithol, byth yn strwythur gwbl statig nac yn olyniad o gyflyrau synchronig, oherwydd mae'r ieithoedd yn byw mewn cnawd dynol, mae rhaid iddynt gael eu caffael a'u hanghofio gan bob cenhedlaeth sy'n mynd heibio ac nid ydynt byth yn llonydd, nac yn gyfyngedig i'w strwythur mewnol. Dyma ble mae dulliau modern, fel swyddogaethiaeth (functionalism) neu ieithyddiaeth wybyddol Evans and Green (2009) yn dod i mewn i'r darlun, ynghyd â seicoieithyddiaeth ddatblygiadol, trwy ddod â'r ffocws i gaffaeliad a defnydd yr iaith a'i berthynas â'r corff, yn hytrach na'i strwythur. Dadleua Bybee (1999) y gall ieithyddiaeth "ddefnydd-seilig" gynhyrchu modelau ffurfiol, ond gyda thro. Trwy ddatgan bod y gymhwysedd yn dod fel rhan o ddefnydd, bron fel priodoledd

allddodol¹, a'r defnydd hwn o'r iaith yn weithgaredd cymdeithasol, corfforol a gwybyddol yn bennaf, mae'r paradigm newydd hwn yn dod ag ystyriaethau newydd i'r golwg. Lle mae cynhyrchioliaeth yn gweld perfformiad fel gwireddu strwythurau cynhenid yr ymenydd gan roi blaenoriaeth i'r strwythur dros unrhyw beth ieithyddol, mae dulliau sy'n seiliedig ar ddefnydd yn ystyried strwythurau fel cyffredinoliadau a wneir gan yr ymennydd a ddysg iaith. Mae'r safbwynt hwn yn mynd y tu hwnt i wrthdroi blaenoriaeth. Trwy bwysleisio bod prosesau gwybyddol bob amser â rhyw radd o ddibyniaeth ar brosesau corffored, synhwyrdd-weithredol, mae'r safbwynt hwn hefyd yn torri deuoliaeth meddwl-corff Descartes (Varela, Thompson and Rosch 1991) yn ogystal â deuoliaeth gymhwysedd-perfformiad Chomsky. Mewn geiriau syml, mae popeth yn yr ymennydd wedi'i gysylltu (neu'n dod i fod wedi'i gysylltu yn y pen draw) yn seiliedig ar ddefnydd, ac mae strwythurau bob amser yn dod a posteriori.

Mae'r datblygiadau hyn mewn ieithyddiaeth ynddi hefyd yn cael eu cefnogi gan ddatblygiad diweddar mewn niwroleg. Ers eu darganfyddiad gan Vernon Mountcastle yn y 1950au, bu dadlau ai yw'r colofnau cortigol sy'n strwythuro'r deunydd llwyd yn y neocortecs yn chwarae rôl fel uned fodiwlwr cyfrifiant (Horton and Adams 2005). Rhagdybiaeth y mil ymennydd (Thousand Brains Hypothesis) (Hawkins, Ahmad and Cui 2017; Hawkins and Dawkins 2021) yw'r iteriad diweddaraf o'r syniad hwn. Mae'n cynnig model ar sut y gall y pensaernïaeth unigryw hon, trwy fecanweithiau pleidleisio, fapio ysgogiadau synhwyrdd-weithredol yn raddol tuag at ac oddi wrth wahanol raddau o haniaethu. Ac ati i fireinio cynrychioliad unedig o'r byd, ac felly ymgysylltu'n well â'r byd mewn dolen adborth barhaus. Mae hyn yn cynhyrchu dadl gymhellol ar sut y gall meddwl haniaethol ac iaith ddod i'r amlwg yn raddol o ryngweithiadau synhwyrdd-weithredol (Constantinescu, O'Reilly and Behrens 2016), pan fo genynnau Gramadeg Cyffredinol Chomsky yn dal i aros i gael eu canfod yn unman.

Goblygiadau ar gyfer Profi Iaith

Ar y pwynt hwn, rhaid egluro'r cyfatebiaeth rhwng paradigm profion y CEFR a ieithyddiaeth ffurfiol, oherwydd ym mharadigm y CEFR, mewn ffordd, rydym yn mesur perfformiad i ddidoli cymhwysedd, felly nid yw'r linc rhwng rheini yn cael ei gwadu. Ond mae'r feirniadaeth epistemolegol o'r ymchwil am gynhwysedd fel un sy'n tanseilio dealltwriaeth o ddynmeg y broses caffael yn dal i sefyll. Os oes gennym ddiddordeb yn y broses caffael a'i ddynmeg, mae cynrychioliad cyflawn, statig o'r sgiliau yn wrthgynhyrchiol. Ymhellach,

¹Emerging property

os nad yw'r cymhwysedd yn bodoli'n annibynnol o'r perfformiad, a ellid didoli'r sgiliau o'r wybodaeth ei hun? Dyma y mae'n ymddangos bod ieithyddiaeth swyddogaethol yn dadlau drosto.

Os yw popeth yn gysylltiedig, os yw popeth yn un (er nad yw un yn bopeth), hynny yw, os yw mwy o ymarfer yn arwain at well sgiliau ymarferol, neu berfformiad, y sy'n arwain at well gwybodaeth ddamcaniaethol, neu gymhwysedd, yna, mewn theori, gallai perfformiad gael ei fesur trwy unrhyw luniad sy'n disgrifio cymhwysedd, megis gwybodaeth eirfa. Mae geirfa yn arbennig o ddiddorol gan fod ei chaffael yn broses ddisgret, ond sy byth yn gorffen yn ystod taith dysgu iaith. Cyhoeddodd Eun Hee Jeon and Yo In'nami (2022) gyfres o meta-ddadansoddiadau ar gydberthnasau'r gwahanol sgiliau ymarferol a ddiffinnir gan y CEFR, i gyd yn pwyntio tuag at y cyfeiriad hwn, gyda gwybodaeth eirfa yn cael ei dyfynnu fel cydberthynas gref ar gyfer rhuglder mewn gwrando (In'nami et al. 2022), siarad (Jeon, In'nami and Koizumi 2022), darllen (Jeon and Yamashita 2022) ac ysgrifennu (Kojima, In'nami and Kaneta 2022). Sylwer fodd bynnag nad yw hyn yn golygu bod gwybodaeth eirfa yn achosi rhuglder, er ei bod yn cyfrannu ato i'r graddau na ddaw rhuglder heb lefel uwch o wybodaeth eirfa. Mae'r rhagdybiaeth sylfaenol hon yn agor y drws ar brofion cyflym â chynllun isel, cost isel, sy'n hygyrch i IAC ac a all fod yn fwy graddadwy a chymwys mewn llawer o feysydd, o hunanasesu, i ddatblygu systemau olrhain dysgu iaith awtomatig fel y grybwyllwyd yn y cyflwyniad. Yn nodedig, yng nghyd-destun IAC, y gall rhai eu galw'n "ieithoedd llafar", mae'r syniad bod lefel eirfa uwch yn gysylltiedig â sgil ymarferol yn dod yn fwy tebygol fyth, oherwydd mai'r ffordd bennaf o gael mynediad at wybodaeth yw "defnydd mwy integredig" (nid yw rhywun yn dysgu Rapa Nui yn y llyfrau). Yn y ffordd hon, gall rhywun hyd yn oed osod y gall profi geirfa ddod yn fwyfwy perthnasol wrth i lai o adnoddau ysgrifenedig a digidol fod ar gael i iaith benodol.

Goblygiad olaf y golwg prif-egwyddor a chysylltiadol hwn o gaffael iaith yw absenoldeb gwahaniaeth ymarferol rhwng y ffordd y ceir cymhwysedd mewn iaith gyntaf (L1) neu ail iaith (L2), hynny yw, trwy ddefnydd. Unwaith yr adeiladir y cylchedwaith sy'n gyfrifol am ddefnydd unrhyw iaith rhwng oedran 1 a 6, naill ai trwy addysg uniaith (gan gynnwys iaith arwyddion) neu amlieithog, mae'r ffordd y ceir geiriau newydd yn gyson ar draws yr ieithoedd a ddysgir yn ddiweddarach yn y fywyd. Os darganfyddir gair neu nodwedd trwy ddefnydd integredig ac mae'r darn gwybodaeth yn yr ymennydd yn tarddu o brofiad synhwyraidd sy'n

bresennol yn ystod caffael y term, ac os dysgir gair yn L2 fel cyfieithiad gair yn L1, bydd ei gynrychioliad yn yr ymennydd yn tarddu o'r gair L1 fel ei gyfystyr o fewn "cofrestr" arall sef rhwydwaith yr L2. Mae'r ddau scenario yn awgrymu ffurfiant gwybodaeth o'r cyd-destun defnydd ond heb wahaniaeth mewn statws rhwng rhwydweithiau L1 ac L2. Gellir dysgu gair yn L2 fel cynnyrch profiad integredig, a gellir dysgu ei gyfwerth L1 yn ddiweddarach fel "cyfystyr mewn gofod arall". Fel rhywun a ddysgodd am ôl-ledaeniad, *backpropagation*, yn Saesneg yn gyntaf, fy nhrydedd iaith, gallaf sicrhau'r darllenydd fy mod yn dal angen meddwl am y gair Saesneg cyn dod o hyd i'w gyfieithiadau yn ystod sgwrs yn Ffrangeg neu Lydaweg, a heb syniad am sut i siarad am hynna yn Gymraeg. Unwaith eto, mae'r gyfatebiaeth hon rhwng L1 ac L2 yn gyfleus yng nghyd-destun IAC, oherwydd bod yr ieithoedd hyn yn aml yn yr amrywiaeth isel mewn rhanbarthau deuglosig, lle mae'r lluniad o siaradwr brodorol a'r llinell rhwng L1 ac L2 yn aml yn aneglur.

2.1.3 Topograffi y Profion Geirfa

Dangoswyd yn aml bod dewis procsiau yn ofalyn yn gallu rhoi darlun dibynadwy o brosesau cymhleth sy'n cael eu mesur. Mae economegwyr wedi dangos er enghraifft sut y gall mesur golau nos o'r gofod wasanaethu fel dangosydd twf CDG² dibynadwy mewn gwledydd lle gall ystadegau swyddogol fod yn brin o ran ansawdd neu onestrwydd (Henderson and Storeygard 2009), hyd yn oed heb ddarparu rheswm pam y gallai hyn weithio. Dychmygodd ieithyddion lawer o ffyrdd i ddiffinio a mesur gwybodaeth eirfa, wrth iddynt ddeall a dangos y gydberthynas gref yr oedd ganddi â rhannau eraill hyfedredd iaith. Bydd y rhan olaf hon o hanner cyntaf yr adolygiad llenyddiaeth yn rhoi trosolwg o'r gwahanol ffyrdd y ceisiodd ieithyddion fesur geirfa hyd yn hyn.

Profion Geirfa Gynhyrchiol

Y ffyrdd mwyaf integredig o brofi geirfa yw gofyn i'r rheini sy'n sefyll y prawf roi cyfystyr rhyw eiriau, gan asesu felly sgiliau geirfa gynhyrchiol, nid yn unig o'r geiriau a ellir eu hadnabod a'u deall, ond hefyd eu hadfer o'u hystyr yn unig. Mae'n un o'r strategaethau a ddefnyddir i fesur y "mynegai geirfa" (vocabulary index), sy'n cael ei gyfuno â thri mynegai arall i gyfrifo IQ y person sy'n sefyll y profion yn sgraddfeydd deallusrwydd oedolion a phlant Wechsler (WISC neu WISA) (Wechsler 2025).

²Cynnyrch Domestig Gros, sef GDP

Profion Geirfa Dderbyniol

Yn ail y ceir cyfres o brofion sy'n anelu at fesur sgiliau geirfa dderbyniol, y geiriau y gellir eu cysylltu â'u hystyr gan y rheini sy'n sefyll y prawf. Y mwyaf defnyddiol o'r rheini yw'r Prawf Lefelau Geirfa (VLT, am *Vocabulary Level Test*), a ddatblygwyd yn y 1980au gan Nation (1990) (gweler Kremmel and Schmitt 2017 am fwy o fanylion am ei weithrediad, datblygiad hanesyddol a chymhwysiadau). Dyluniwyd y prawf hwn ar gyfer defnydd eang mewn ysgolion fel prawf lleoli myfyrwyr. Mae VLT yn addasol yn rannol hefyd, gan ei fod yn profi'r sgiliau i gysylltu termau sy'n gysylltiedig o ran ystyr o wahanol ystodau amllder. Dyluniad prawf geirfa dderbyniol diddorol arall yw'r brawf geirfa darluniol Peabody (L. Dunn and D. Dunn 2025). Gan ei fod yn seiliedig ar luniau yn hytrach na geiriau ysgrifenedig, mae'n caniatáu profi plant na allent fel arall ddarllen y geiriau sy'n cael eu hasesu. Gallai'r dull sy'n seiliedig ar luniau hwn ymddangos fel pe bai'n gwneud y dyluniad profi hwn yn addas delfrydol i'w raddio yn blaenorol, ac felly'n addas ar gyfer safon gyffredinol y gellid ei chymhwyso hyd yn oed mewn amgylcheddau lle nad yw llythrennedd yn eang. Fodd bynnag, efallai bod y syniad hwn yn dda ar yr wyneb yn unig, gan fod y graddnodi ar gyfer mapio lluniau-geiriau wedi digwydd mewn gwlad sy'n siarad Saesneg, a gall y geiriau a ddefnyddir i ddisgrifio sefyllfaoedd tebyg amrywio'n fawr rhwng gwahanol ofodau ieithyddol. Dyma ddysgodd Kartushina et al. (2022) y ffordd anodd wrth iddynt geisio cyfieithu'r prawf yn Rwseg ar gyfer plant cyn-ysgol, gan ddangos braidd yn ddamweiniol efallai mai prawf Peabody yw un o'r profion geirfa anoddaf i'w addasu i ieithoedd eraill, hyd yn oed rhai a siaredir mewn cymdeithasau sy'n braidd yn WEIRD fel Rwsia.

Profion Geirfa Adnabod

Yn olaf, y teulu symlaf o brofion geirfa yw'r profion geirfa adnabod, a elwir weithiau'n "brofion geirfa syml". Maent yn mesur y gallu i adnabod presenoldeb gair cywir yn unig, heb fynnu cyfiawnhad o ddealltwriaeth bellach o ystyr y gair. Am drosolwg ac asesiad o wahanol ddyluniadau, gweler P. Meara 1994. Y dyluniad mwyaf llwyddiannus o'r teulu profi geirfa syml hwn yw'r prawf geirfa seiliedig ar dasg penderfynu geirfaol (LDT am *Lexical Decision Task*), rhoddwyd llawer o enwau eraill iddynt megis profion geirfa "Ie/Na" neu "ddeuaidd", ond mae'r cyfan yn dilyn yr un egwyddor; cyflwynir dilyniant o eitemau i'w profi, naill ai geiriau go iawn neu ffug-eiriau (Paul Meara 2012) i'r rheini sy'n sefyll y prawf. Yna, gofynnir iddynt yn systematig a ydynt yn credu bod yr eitem dan sylw yn perthyn i eirfa'r

iaith. Daw'r canlyniadau ar ffurf cyfuniad o'r pedwar allbwn a ddiffinnir gan fatrics dryswch³ (cyfuniad o atebion Cywir neu Anghywir a Phositif neu Negatif), sef trawiadau (PC), methu (NA), larwm ffug (PA) a gwrthod cywir (NA) ac mae gwahanol fethodolegau wedi cael eu cynnig i drin y canlyniadau, o dynnu canran yr atebion anghywir o ganran yr atebion cywir, hyd at gymhwyso systemau mwy cymhleth o Theori Canfod Signalau (TCS) (Huibregtse, Admiraal and Paul Meara 2002).

Mae llawer o brofion o'r fath wedi cael eu hadeiladu hyd yn hyn yn cynnwys o leiaf un fersiwn ar-lein, ar gael mewn sawl iaith, sef Saesneg, Iseldireg ac Almaeneg (Lemhöfer and Broersma 2012). Dangosodd y papur hwn ganlyniadau calonogol, gan ddangos gydberthynas gref o'r canlyniad geirfa gyda phrofion traddodiadol eraill, gan gefnogi felly'r syniad y gellir mesur rhuglder yn effeithiol trwy brofi geirfa. Mae prawf arall sydd, ar ei semlant, wedi cael ei greu ar gyfer Croateg (Srce 2025), er nad yw mwy o wybodaeth ar gael eto. Ac mae hyn ochr yn ochr â'r systemau niferus a ddatblygwyd gan Meara dros y blynyddoedd^P. Meara 1994. Prif gyfyngiad y systemau hyn yw'r ffaith bod eu heitemau'n gyfyngedig a statig, felly ni chânt byth eu dylunio ar gyfer defnydd ailadroddus, a fyddai'n helpu mesur dynameg caffael geirfa. Mae hon yn broblem i'w datrys, oherwydd prif ddiddordeb prawf minimaliaidd yw caniatáu profion cylchol. Ar y llaw arall, ymddengys mai hwy yw'r profion byrraf i sefyll, yr hyn sy'n fantais pwysig.

2.1.4 Perthnasedd a Chyfyngiadau Profion Geirfa

Mae'r holl brofion geirfa a gyflwynwyd uchod wedi mwynhau rhyw raddau o lwyddiant masnachol neu academiaidd oherwydd eu dibynadwyedd wrth ddal agweddau gwahanol ar gaffael geirfa. Rydym eisoes wedi esbonio'r rhesymau pam y dylai hyn fod felly yn adran 2.1.2. Os derbynnir bod unrhyw is-adeiledd o hyfedredd yn gysylltiedig yn yr ymennydd mewn ffordd a ddiffinnir gan ddefnydd, bod "popeth yn un", yna mae'r un rhesymeg yn berthnasol i eirfa. Mae adnabod yn dod fel y cam cyntaf yn ystod caffael geirfa. Heb adnabod, mae datblygiad ymhellach tuag at ddefnydd mwy itegredig yn amhosib. Mae'r holl deuluoedd profi hyn yn mesur gwahanol gamau yr un broses o gaffael geirfa. Nid yw mesur golau nos yn mesur y lluniad "defnydd trydan nos-amser wedi'i canolbwyntio ar gyfer goleuadau stryd tiriogaeth" yn unig, ond, fel y dangosodd yr ystadegau, gellir ei ddefnyddio fel dangosydd CMG, y sydd ei hun yn ddangosydd iechyd economaidd gwledydd. Yr un peth sy'n wir am

³Gweld Confusion Matrix

y profion geirfa hyn, mae'r cyfan yn lluniadau sy'n mesur camau gwahanol yr un ffenomen caffael geirfa, sy'n rhan annatod o gaffael iaith.

Y brif wahaniaethau rhwng y profion hyn yw pa mor adnodd-ddwys ydynt a pha mor integredig yw'r lluniadau y maent yn eu mesur. Mae dangosyddion syml fel adnabod geirfa yn unig â gwendidau a gall fod yn destun twyllo neu lawdriniad. Honnwyd bod y mynegai enwog "Big Mac" yr Economist ar gyfer chwyddiant yn darged ymgaisiau llawdriniadau gan lywodraeth yr Ariannin yn 2011 (Politi 2011), am yr union reswm hwn. Yn yr un modd, yr hawsaf yw caffael y sgil y tu ôl y lluniad a ddefnyddir fel dangosydd, y mwyaf tebygol y bydd yn dod yn destun ymdrechion i'w dylanwadu. Ond nid yw hyn yn golygu nad oes gan y lluniad unrhyw werth, yn wir, mae lefelau golau nos a phrisiau Big Mac yn dal i gael eu defnyddio heddiw, ond mewn meysydd a mewn cyd-destyn sy'n berthnasol i'w cymhlethdod. Yr un peth sy'n wir am seicometreg. Mewn cyd-destun profion geirfa, mae'r gofyniad dylunio cymhleth sy'n gorfod defnyddio llawer o adnoddau yn y Peabody Picture Vocabulary Test yn ei gwneud yn anghrheidiol cael defnydd masnachol i gefnogi ei ddatblygiad cymhleth. Mae'r profion eraill, symlach yn cyflawni dim ond llwyddiant academiaidd oherwydd eu bod mor syml i'w rhoi ar waith fel na fyddant byth angen fasnacheiddio, sy'n cyfyngu ar eu potensial graddio ac yn ei dro eu datblygiad. Serch hynny, maent i gyd yr un mor ddefnyddiol wrth fesur eu camau priodol o gaffael geirfa, ac ar gyfer amcan y gwaith presennol, wrth fesur dynameg y proses caffael iaith.

2.1.5 Casgliad

Yng nghyd-destun profi awtomatig ac addasol gyda'r nod o olrhain caffael sgiliau iaith, mae manteision profion geirfa yn amlwg yn uwch na dulliau eraill. Yn eu plith, mae dyluniadau profion adnabod geirfa symlach yn wirioneddol ddisgleirio, yn enwedig wrth ystyried y broblem a osodir gan IAC. Mae profion geirfa LDT yn symlach i'w gweinyddu mewn ffordd gwbl awtomatig, ac maent yn haws eu trosglwyddo i IAC oherwydd gellir eu deillio o restr syml o gofnodion geiriadur. Eto, erys heriau sylweddol yn parhau cyn galluogi gweithrediad eang o brawf geirfa LDT. Y ffactor cyfyngu pennaf yw nifer yr eitemau a gynigir yn y profion fel LexTALE, pan oedd yn rhaid dewis geiriau go iawn a ffug-eiriau o set fwy mewn astudiaeth ragbaratol (Lemhöfer and Broersma 2012). Os yw prawf geirfa LDT i gael ei ddefnyddio mewn ffordd ddychweliadol, i olrhain cynnydd geirfa trwy amser, rhaid bod digon o eitemau ar gael, efallai'n cynnwys holl eirfa iaith neu o leiaf cyfran sylweddol ohoni. Ond felly, mae'r cwestiwn o raddnodi'r eitemau'n yn dod yn bwysig. Ni ellir ystyried defnyddio'r

astudiaeth ragbaratol a wnaed ar gyfer dewis eitemau yn LexTALE i greu digon o eitemau i ganiatáu profi dychweliadol dibynadwy, hyd yn oed ar gyfer iaith â chyfoeth anhygoel o adnoddau fel Saesneg, heb sôn am ieithoedd adnoddau cyfyngedig. Byddai datrys y broblem hon o raddnodi'r eitemau yn agor y drws i raddio y math o bofion geirfa'n fertigol (caniatáu profi yn ddychweliadol yr un bobl) ac yn llorweddol (caniatáu cludo'r prawf i lawer o wahanol ieithoedd). Bydd yr adran nesaf yn cael ei canolbwyntio i ddod o hyd i ateb o'r fath.

2.2 Olrhain Gwybodaeth

I baraffrasu P. Meara (1994), gall llawer o dasgau asesu fod yn ffyrdd dilys o asesu sgiliau adnabod geirfa, cyn belled â bod y dull priodol o ddadansoddi'r canlyniadau'n cael ei ddarparu. Mae'r adran hon yn canolbwyntio ar y broblem darparu hon. Tasg gymhleth yw mesur nodweddion cudd o ymatebion eitemau, pa un a adwaenir fel Olrhain Gwybodaeth (OG) (Shen et al. 2024). Cysyniad sylfaenol mewn Profi Addasol Cyfrifiadurol (PAC) yw cwestion dewis model OG dilys. Mae rhan o'r cymhlethdod hwn yn dibynnu ar y rhagdybiaethau y mae rhywun yn eu gwneud ar y nodweddion cudd hyn, ai ydynt yn lluniad parhaus neu'n set o sgiliau arwahanol, sy'n cyfuno gyda'i gilydd mewn gofod gwybodaeth aml dimensiwn, ac os felly, pa sgiliau sy'n dibynnu ar ba rai eraill? Gellir deffinio'r dimensiynau hyn a'r perthnasedd rhyngddynt â llaw neu yn seiliedig ar ddata, gan ddefnyddio technegau Bayesaidd neu rhywdweithiau niwral a DD. Gall rhagdybiaethau eraill gynnwys dylanwadau y broses brofi ar y broses ddysgu, ac yn yr achos hwnnw gall rhywun ystyried hanner-oes atgofion newydd a ffurfiwyd yn ystod rowndiau asesu blaenorol fel gan systemau “flash-cards”, sydd yn seiliedig ar ailadroddiad bylchog (spaced repetition). Yn sylfaenol, mae'r dewis cymhleth hwn o'r model yn arbitriad rhwng cywirdeb a dealladwyedd (Pelánek 2025). Efallai fod modelau mwy ansoddol yn briodol i hysbysu argymhellion o adnoddau dysgu, ond cyflwyno canlyniadau prawf annibynnol fel fector rhuglder a fyddai yn galetach i'w ddehongli nâ sgôr glasurrol. Problem arall yw, in fine, bod angen hyfforddiant aruthrol ar fodelau ansoddol i weithio, maent hywthau yn fodelau meintiol mewn fordd.

Gan fod y traethawd hwn yn canolbwyntio'n bennaf ar brofi twf geirfaol, mae mynegai unddimensiwn yn ymddangos yn briodolach. Ymhellach, byddai graddnodi paradigm aml dimensiwn yn gofyn am swm mawr o ddata neu adnoddau fel amser ac arbenigedd, nad ydynt ar gael ar gyfer IAC. Bydd diwedd y bennod hon yn gosod y sylfaen ddamcaniaethol ar gyfer y dehongliad meintiol hwn o ganlyniadau prawf geirfa LDT.

2.2.1 Gallu Damcaniaethol Prawf Unddimensiwn Diswn

Nod model olrhain gwybodaeth mewn PAC yw rhagfynegi peth byddai canlyniad profiad eitemau'r prawf er mwyn dewis y rhai y mae eu hatebion yn fwyaf ansicr, gan bob rhagfynegiad yn seiliedig ar ganlyniadau blaenorol. Yn jargon theori gwybodaeth, gelwir hyn mwyafu'r entropi, sy'n mwyafu'r elw gwybodaeth trwy leihau ansicrwydd y model. Gan dynnu o Shannon (1948), gall rhywun ddefinio'n ddamcaniaethol y gallu diamod mewn prawf deuaidd "diswn", cyn ei addasu i amgylchedd swnllyd. Mewn graddfa syml, unddimensiwn, gellir cyflawni dod o hyd i'r man ansicrwydd uchaf hwn gyda'r algorithm chwilio deuaidd. Mewn restr o eitemau wedi'u trefnu yn ôl anhawster, cymryd eitem yn y canol, ailadroddwch y broses gydag ail hanner y rhestr wreiddiol os yw'r ateb yn gywir (sef, bu'r eitem yn rhy awdd), fel arall, gyda'r hanner cyntaf. Ailadroddwch y broses nes bod y rhestr yn un eitem o hyd. Mae gan yr algorithm hwn gymhlethdod amser o $\theta(\log n)$, sy'n golygu bod angen $\log_2(n)$ cam ar gyfer n nifer o eitemau i gyrraedd yr eitem olaf. Dyma 10 eitem sydd angen eu profi ar gyfer graddfa sy'n cynnwys 1 024 o eitemau, 11 ar gyfer 2 048 eitem, 12 ar gyfer 4 096 ac yn y blaen...

Gan dybio y gellid trefnu'r holl eiriau mewn geiriadur (bach) o 30 000 o eiriau yn ôl eu "hanhawster", a bod rhaid i hanner eitemau prawf fod yn ffug-eiriau i atal twyllo, byddai prawf seiledig ar yr algorithm hwn yn dod o hyd i lefel gyfredol y person sy'n sefyll y prawf mewn dim ond 30 rownd o brofi, i'w gymharu â'r 60 eitem (40 geirau gwir, 20 ffug) a ddefnyddwyd gan brawf fel LexTALE (Lemhöfer and Broersma 2012). Hyd yn oed os ydym yn ystyried yr angen ar gyfer cywiro gwallau, bydd cyfanswm nifer y camau sydd eu hangen yn parhau i fod yn gyfrannol i'r dilyniant logarithmig hwn. Mae gan y gosodiad hwn gyfyngiadau amlwg ac elir i'r afael ganddynt yn yr isadran ganlynol, ond mae'n dod â dirnadiaeth diddorol ynghylch y broblem gwneud profion graddadwy. Yn bennaf, mae'n bosibl profi nifer mawr iawn o eitemau mewn ffordd effeithlon o ran amser, y peth sy'n agor y drws i ddefnyddio holl eirfa iaith, neu o leuaf rhan fawr ohoni, fel eitemau profi, yn hytrach na rhestr fer o eiriau detholedig. Mae'r posibilrwydd hwn yn ei dro yn agor y drws i profion unigryw, lle mae'r siawns o fynd ddwywaith trwy'r un geiriau mewn dwy brawf bron yn amhosibl. Mae hyn yn datgloi problem raddio fertigol a amlygwyd yn gynharach yn y bennod hon.

2.2.2 Y System Graddio Elo

Graddio Elo a'r Model Rasch

Cyfngiad amlwg cyntaf y model a gynigwyd yn flaenorol yw calibrïad yr eitemau. Ni all rhywun gael lefel anhawster rhyw air yn uniongyrchol o eiriaduron, a gall y drefn y mae dysgwyr yn caffael geiriau amrywio'n fawr yn dibynnu ar wahanol ffactorau. Mae'r rhan fwyaf o brofion geirfa yn mynd o gwmpas y broblem hon drwy grwpio'r eitemau yn ôl ystodau amllder (Nation 1990; P. Meara 1994; Dudley, Marsden and Bovolenta 2024). Fodd bynnag, nid oes gan y rhan fwyaf o IAC adnoddau o'r fath ar gael iddynt, neu os oes. Am y rheswm hwn, rydym yn cynnig y dylai graddio anhawster yr eitemau geiriau gael ei ddiweddar'u'n uniongyrchol yn seiliedig ar ganlyniadau cynt y prawf.

Mewn profion safonol, cyflawnir y calibrïad hwn o anhawster yr eitemau gan Theori Ymateb Eitemau (TYE), sy'n set o fodelau a ddeilliwyd o'r model Rasch (Rasch 1980). Ailddargau fuwyd y mathemateg y tu ôl i'r model Rasch sawl gwaith, gan gynnwys y tu allan i'r byd seicometrig, fel mewn gwyddbwyll gyda'r system graddio Elo (Elo 1961; Elo 1986). Cyflwynir yr hafaliadau allweddol ar gyfer y modelau hyn isod.

$$P(X_{AB} = 1) = \frac{1}{1 + e^{R_b - R_a}}$$

Figure 2.1 – Fformiwla Rasch

$$P(X_{AB} = 1) = \frac{1}{1 + 10^{\frac{R_b - R_a}{400}}}$$

Figure 2.2 – System graddio Elo

Yn y system graddio Elo, $P(X_{AB} = 1)$ sydd y tebygolrwydd y bydd chwaraewr gwyddbwyll A o radd R_a yn ennill trwy siachmat yn erbyn chwaraewr B o raddio R_b . Yn y model Rasch, $P(X_{AB} = 1)$ sydd y tebygolrwydd y bydd rhyw berson A gan radd R_a wrth sefyll prawf yn ateb yn gywir eitem o radd anhawster R_b . Gan fod y ddau yn dilyn dilyniant logarithmig, gellir trosi unrhyw “radd Rasch” i radd Elo drwy ei luosi â $400/\ln(10)$ (neu $400 \times \log_{10}(e)$) a gwrthdroi'r enwebwr a'r enwadur i fynd o Elo i Rasch. Roedd y gwahaniaeth yn sylfaen y logarithm a'r ychwanegiad o “ffactor taeniad” o 400 mewn gwyddbwyll yn golygu cynyddu darllenadwyedd a dehongliadwyedd, wrth gyfateb i systemau graddio a ddefnyddiwyd yn flaenorol yn y byd gwyddbwyll. Mae gwahaniaeth o 400 mewn graddio Elo yn golygu siawns buddugoliaeth o 1 : 11 a 10 : 11, sy'n fwy dehongladwy na gwahaniaeth o 1 pwynt yn golygu dosbarthiad siansiau o 0.2689 : 0.7311 ac 0.7311 : 0.2689.

Yn ymarferol, mae'r prif wahaniaeth rhwng y ddwy system yn gorwedd yn fwy yn y mecanweithiau diweddarau. Gan fod TYE wedi'i ddatblygu ar gyfer profion statig, heb nodweddion ymaddasu amser real, a gan atebion i bob eitem am bob myfyriwr, mae'n dibynnu ar dechnegau sy'n ddwys yn gyfrifiadol, nad ydynt yn addas iawn at ddiben PAC. Ei system ddiweddarau syml yw'r rheswm pam mae'r system graddio Elo wedi bod yn denu mwy o sylw yn y gymuned AIED dros y blynyddoedd. Mae Pelánek 2016 yn crybwyll nifer o integreiddiadau llwyddiannus o'r model hwn mewn gosodiadau addysgol ymaddasol, er byth ar gyfer profion annibynnol. Mae'r un erthygl hefyd yn cyflwyno mecanweithiau diweddarau amrywiol sy'n ystyried gwahanol tybiaethau, megis cywiriad wrth strategaethau twyllo neu hanner oes cof tymor byr a chanolig. Rhoddir diweddariad graddio Elo gan y fformiwla ganlynol.

$$R'_A = R_A + K \times (S - P(S)) \quad (2.1)$$

Mae'r sgôr gwirioneddol (1 neu 0) S yn cael ei dynnu yr rhagfynegiad $P(S)$ yn seiliedig ar y gwahaniaeth sgôr a roddir yn y ffig. 2.2. Os yw canlyniad yn sicr (sef tua siansiau 1:0 gan mwy na gwahaniaeth graddio o 800) ac mae'r canlyniad yn dilyn y rhagfynegiad, bydd y gwerth hwn yn agos at sero a bydd y newid mewn graddio yn agos at 0. Os yw'r gwrthwyneb yn digwydd, mae'r sgôr yn cynyddu gan werth sy'n agos at K , a enwir yn "ffactor K ", gwerth sy'n debyg i'r gyfradd ddysgu yn y byd DL. Gall y gwerth hwn amrywio yn dibynnu ar weithrediadau'r system graddio, ond yn aml mae oddeutu 20 yn y byd gwyddbwyll. Weithiau, defnyddir "ffwythiant ansicrwydd" i newid y gyfradd ddiweddarau yn raddol yn seiliedig ar nifer y diweddariadau cynt (cf. hafaliad 2.2.2).

Cywiro Gwallau

Mae'r Holiadau Amlddewis (HAdd) yn defnyddio tri chategori o gydran, cwestionau neu ymholiadau, am *query* yn Saeseg, *allweddi*, sef yr atebion gwirion, a *llithiau*, a elwir *distractors* yn Saesneg, sef yr atebion anghywir a ddefnyddir i lithio, twyllo'r myfyrwyr. Yn sylfaenol, mae profion geirfa adnabod yn ffurf o HAdd, gydag ymholiad unigryw ar gyfer y prawf cyfan, a'r geiriau go iawn fel allweddi a'r ffug-eiriau fel llithiau. Cydnabyddir y gall fod rhesymau gwahanol pam y gall sefyllwyr prawf ddewis atebion cywir neu anghywir. Yr

un mwyaf amlwg yw bod sefyllwyr prawf yn adnabod yr allweddi ac yn anwybyddu'r llithiau. Ond rhaid ystyried tri cwrs gweithredu arall.

1. Mae'r sefyllwyr prawf yn gwybod yr ateb ond yn dewis ateb anghywir ar gam (e.e. drwy ateb yn rhy gyflym a sylwi ar y camgymeriad yn rhy hwyr).
2. Nid yw'r sefyllwyr prawf yn gwybod yr ateb cywir, ac mae'n ateb yn briodol drwy siawns bur.
3. Nid yw graddio'r eitem yn cyfateb i'w lefel anhawster gwirioneddol oherwydd nad yw'r calibriad ar ben (ac yn wir, ni all fod byth yn bertffaith gan nifer mawr o eitemau).

Deallir bod yr effeithiau hyn yn ychwanegu sŵn at y system ac y dylai ddodi i'r prawf fwy o afreidrwydd i gywirio'r effeithiau hyn. Deallir pe bai ateb yn cael ei roi am reswm da o leuaf fwy na hanner yr amser, byddai graddio'r sefyllwyr prawf yn dal i gydgyfeirio tuag at ei werth go iawn, er yn arafach. Ond beth bynnag, ni ddylai nifer yr eitemau a brofwyd mewn sesiwn prawf gael ei wneud mor fyr ag sy'n bosibl yn ddamcaniaethol, ond cymryd y sŵn hwn i ystyriaeth. Unwaith eto, mae'r system graddio Elo yn gwneud hyn yn ddi-dor gyda'r "ffwythiant ansicrwydd" a fynegwyd yn barod. Mae Pelánek 2016 yn cynnig y ffwythiant ansicrwydd canlynol i ddiweddarau cyfradd diweddariad (sef gwerth y ffactor K uchod) yn ffwythiant o nifer yr atebion blaenorol.

$$(n) = a/(1 + bn) \quad (2.2)$$

Lle mae a a b yn gysonion positif,⁴ ac n yw nifer yr eitemau a atebwyd yn flaenorol. Defnyddir y rhif sy'n deillio fel y ffactor K sy'n lluosu cywiriad graddio ar ôl pob ateb. Unwaith eto, byddwn yn dychwelyd at yr agwedd hon yn y bennod nesaf.

2.3 Casgliad

Cyflwynodd yr adolygiad llenyddiaeth hwn syniadau o sawl maes a cheisiodd eu trefnu yn gyfanwaith cydlynol. O ddadlau seicoieithyddol sy'n cefnogi'r syniad y gellir defnyddio geirfa fel procsi ar gyfer cymhwysedd iaith gyffredinol, i gynnig model olrhain gwybodaeth

⁴Mae Pelánek (Pelánek 2016) yn cynghori $a = 1$ a $b = 0.005$ pan raddir yr eitemau gan 2.1 sef $a = 921$ a $b = 4.6$ gan "daeniad" y system a ddefnyddir mewn gwyddbwyll.

sy'n optimeiddio'r wybodaeth a enillir gan ganlyniadau prawf deuaidd. Yn y bennod nesaf, byddwn yn rhoi'r darnau hyn gyda'i gilydd i adeiladu prawf geirfa adnabod cyflawn a ffwythiannol.

Pennod 3

Methodoleg

Mae'r bennod hon yn disgrifio sut bu prawf geirfa deuaidd yn cael ei lunio yn seiledig ar y pwyntau a amlygwyd yn yr adolygiad llenyddiaeth. Pan fydd agweddau arbennig i ieithoedd yn cael eu mynegu, fel sut i ffynonellu'r geiriau ar lein o eiriadur Llydaweg, deallir y gellir defnyddio'r un dull i greu test mewn iaith arall, pan ni wnaethpwyd o'r blaen.

3.1 Ffynonellu'r Allweddïau

Ar gyfer y prawf geirfa Llydaweg, casgwyd holl gofnodion geiriadur diachronig Llydaweg Devri.bzh, a chynlluniwyd rheolau i dynnu enwau priod ac ôl-ddeiliaid. Gan fod y geiriadur unieithog Meurgorf¹ yn dosbarthu ei gofnodion yn un o dri chategori: aml, cyffredin a phrin, roedd yn bosibl trefnu'r eitemau mewn pedwar categori, y tri blaenorol, a'r eitemau a oedd yn Devri ond nid yn Meurgorf. Dangosir dosbarthiad yr eitemau yn 3.1. Y rhesymau pam y mae cymaint o eiriau yn ymddangos fel pe baent yn absennol yn Devri, yw bod cofnodion Meurgorf yn cynnwys llawer o enwau priod ac ôl-ddeiliaid, yn ogystal â neologismau wedi'u hadeiladu gydag ôl-ddeiliaid cyffredin. Cyfanswm nifer yr allweddi sydd ar gael ar gyfer y prawf oedd 62,169, y rhoddwyd i hanner ohonynt sgôr anhawster rhwng 1 a 3. Ychwanegwyd y cofnodion o Devri na ddarganfuwyd yn Meurgorf at gategori'r tipau prin.

categoriâu	ym Meurgorf	yn Devri hefyd	yn unig yn Devri
Aml	1 108	946	—
Cyffredin	47 740	26 197	—
Prin	6 867	4 868	—
Cyfanswm	55 715	32 011	30 158

Table 3.1 – Categoriâu o eiriau a dynnwyd o Devri (wedi'u hidlo) a Meurgorf

Yn amlwg, nid yw'r categorïau hyn yn berffaith, ond maent yn dal yn gymorth gwerthfawr i gwestiynau anodd graddnodi. Gallai dulliau eraill o gyrchu tipau ac ystodau gwahanol o amllderau ar gyfer ieithoedd eraill gynnwys cyrchu geiriaduron o feintiau gwahanol, gan fydd

¹Gweler <https://niverel.brezhoneg.bzh/br/meurgorf/>

y geiriau yn y geiriaduriau lleiaf y geiriau amlaf. Mae'r adran ar gychwyn graddio yn dangos sut y defnyddir y graddau cychwynnol hyn. Gellir dod o hyd i'r cod ar gyfer y camau hyn ar GitHub².

3.2 Cynhyrchu'r Llithiau

3.2.1 Hyfforddi'r Model

Am astudiaeth o raddfa'r prosiect hwn, nid yw creu'r geiriau ffug â llaw yn opsiwn. Datblygwyd dulliau gwahanol o greu geiriau ffug yn gyfrifiadurol dros y blynyddoedd, mae'r rhan fwyaf ohonynt yn cadwyni n-gramau a gymerwyd o set ddata hyfforddi o feintiau gwahanol (New et al. 2023; Keuleers and Brysbaert 2010). Fodd bynnag, gan fod rhai ieithoedd yn arddangos nodweddion perthnasoedd pellter hir ffônotechnig, megis cysondeb ynganiad mewn ieithoedd Twrcaidd, ystyriwyd nad oedd modelau seiliedig ar n-gramau yn orau ar gyfer creu geiriau ffug tebygol. Am y rheswm hwn, rhoddwyd blaenoriaeth i ddefnyddio Cof Byr Ystod Hir (neu LSTM am Long Short-Term Memory) (Hochreiter and Schmidhuber 1997). Bydd y dyluniad yn syml i bobl sy'n gyfarwydd â Rhwydweithiau Niwral Ailadroddus (neu RNN am Recurrent neural Network), ond datblygwyd rhai technegau optimeiddio i gynyddu cyflymder yr hyfforddiant. Gan fod y geiriau yn y set ddata hyfforddi (y allweddi o'r adran flaenorol) o hydoedd amrywiol, mae pop "batch" o hyd 1, peth sy'n golygu na allai unrhyw baraleliaeth fod yn bosibl yn ystod hyfforddiant. I osgoi'r broblem hon, cysylltwyd y geiriau mewn cant o linyddau hirach, gyda nodyn llinell newydd a ddefnyddiwyd fel y tocyn arbennig, i ddechrau dilyniant o eiriau, i wahanu pob gair ac i orffen pob dilyniant, gan wneud y dilyniannau hyn yn gryno ac yn ddarllenadwy gan bobl. Cadwyd 15 o'r dilyniannau hyn ar gyfer dilysu a 85 ar gyfer hyfforddiant priodol. Roedd tua 10 dimensiwn mewnosod (i gynrychioli'r cymeriadau) a 180 dimensiwn cudd (i gofio'r patrymau) yn ddigonol i hyfforddi ein model iaith "orthograffig" a allai gynhyrchu geiriau ffug o ansawdd da. Mae rhwng 10 a 20 o epocau yn ddigon i gael entropi croes is na 1.8, a diolch i'r dechneg batio a grybwyllwyd uchod, ni chymryd yr hyfforddiant bron yn 10 i 12 eiliad yr epoc. Nodwch, fodd bynnag, nad oedd entropi croes bop tro is yn cael ei gael yn systematig, hyd yn oed gyda'r un hyperparamedrau, gan fod y proses yn stocastig. Dyma lle mae techneg optimeiddio arall yn dod i'r amlwg, effaith synhwyrol ar gynnydd y ffwythiant colled trwy ail-drefnu'r geiriau mewn

²Ar gyfer cyrchu geiriau Devri a'u hidlo, gweler y llyfr nodiadau Jupyter hwn: <https://github.com/Oktogazh/sudogen/blob/master/1%20Introduction.ipynb>, ar gyfer yr ystod amllderau, gweler y llyfr nodiadau Jupyter arall hwn: <https://github.com/Oktogazh/sudogen/blob/master/locales/br/5%20Initialization.ipynb>

trefn wahanol a gwneud dilyniannau gwahanol o'r un geiriau. Yn ffordd, roedd hyn yn “creu mwy o ddata hyfforddi” lle mai'r unig bwynt cyffredin rhwng y dilyniannau blaenorol a'r newydd oedd strwythur mewnol y geiriau, a byddai'r cysylltiadau rhwng y geiriau ddim yn cael ei hystyried gan fector cudd LSTM. Roedd hyn yn cyfnewid rhai eiriau o'r set dilynol i'r set hyfforddiant, ond gan fod colled ffwythiant y set ddata hyfforddi'n is na chyson ar gyfer pob hyfforddiant, gan ddangos dim arwyddion o “over-fitting” hyd yn oed gyda nifer fawr o epocau, roedd hyn yn cael ei ystyried nad oedd yn broblem.

3.2.2 Cynhyrchu Geiriau Ffug

Unwaith y cafwyd model iaith seiliedig ar gymeriad wedi'i hyfforddi gydag entropi croes boddhaol, mae'n barod i gynhyrchu geiriau newydd. Canfuwyd bod tymheredd o 0.7 yn “bwynt melyn” ar gyfer cydbwysedd da rhwng amrywiaeth a chywirdeb y characterau a gynhyrchir. Cafodd y pwynt melyn hwn ei ganfod trwy edrych ar gyfran y geiriau sy'n dechrau gyda'r llythyren z, yn Llydaweg 1:2000 mathau sy'n dechrau gyda z. Yn amlwg, byddai ieithoedd gwahanol, yn enwedig ieithoedd ag alfabetau eraill angen tymheredd arall.

Gan fod nod y rhwydwaith yw ail-gynhyrchu'r data hyfforddi gyda'r ffyddlondeb fwyaf posibl, bydd yn ceisio cynhyrchu geiriau go iawn. Rhaid hidlo'r geiriau go iawn, a wnaethpwyd mewn dwy ffordd wahanol. Bob tro y cynhyrchwyd gair newydd, pan gynhyrchwyd y character llinell newydd, yna cymharir y llinyn characteriau newydd gyda'r tipau sydd ar gael yn y set ddata hyfforddi, os nad yw yn y set ddata hyfforddi, mae'r gair yn cael ei wirio yn erbyn geiriadur sillafu Hunspell. Dim ond os na chydabyddir y sillafu yna, mai ychwanegir y gair at set o eiriau ffug a gynhyrchir, gyda gradd uchel o sicrwydd bod y gair yn ddibwys. Mae'r cod ar gyfer cynhyrchu'r geiriau ffug ar gael ar GitHub³. Defnyddiwyd y dull hwn i gynhyrchu nifer gyfartal o eiriau ffug â geiriau go iawn.

Gan fod un peth a allai roi geiriau ffug yn ôl yw anghydbwysedd rhai characteriau yn y geiriau, defnyddir y Ffigur 3.2.2 i archwilio dosbarthiad y characteriau trwy'r setiau o eiriau a geiriau ffug. Os yw'r gwerth ar gyfer llythyren benodol yn bositif, mae'n golygu bod y character yn cael ei or-ddatblygu yn y set o eiriau ffug, a'r gwrthwyneb os yw'r gwerth yn mynd yn negyddol. Mae'r characteriau wedi'u trefnu yn ôl amllder, *e* yw'r cymeriad mwyaf cyffredin a *ē* yw'r lleiaf cyffredin yn Llydaweg (a geir unwaith yn unig yn yr eiriau go iawn, ac byth

³Gweler y llyfr nodiadau Jupyter hwn am fanylion: <https://github.com/Oktogazh/sudogen/blob/master/2%20Training.ipynb>

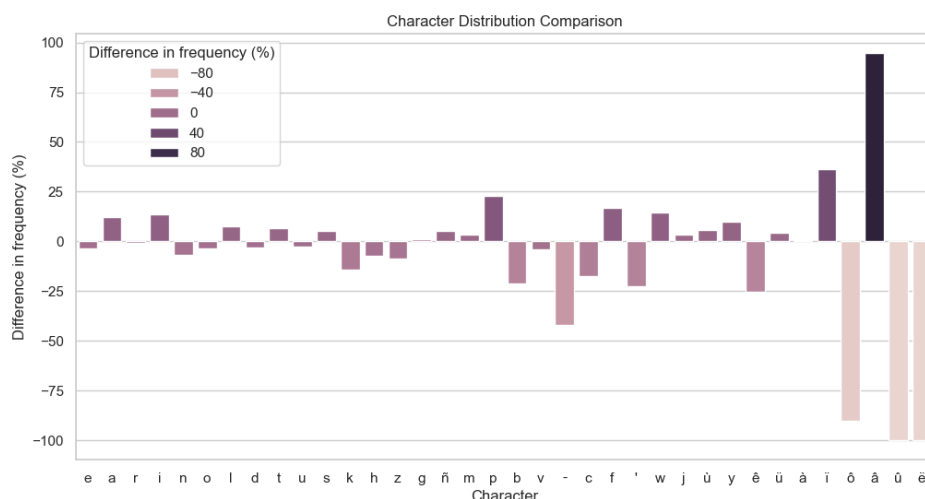


Figure 3.1 – Gwahaniaeth mewn dosbarthiad y characteriau (geiriau ffug / geiriau go iawn)

yn yr eiriau ffug, gan hynny'r gwahaniaeth o 100%). Yn gyffredinol, mae'r' characteriau yn y geiriau ffug a gynhyrchir yn ymddangos yn gydnaws ag eu dosbarthiad y geiriau go iawn.

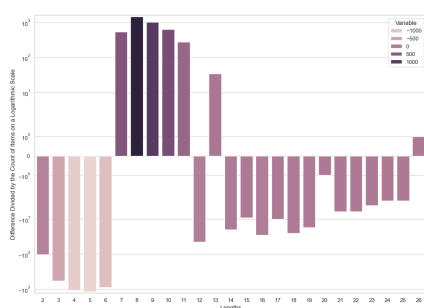


Figure 3.2 – Gwahaniaeth rhwng y cyfiri o eiriau ffug dros eiriau go iawn ar raddfa logariffmig ar gyfer hyd penodol.

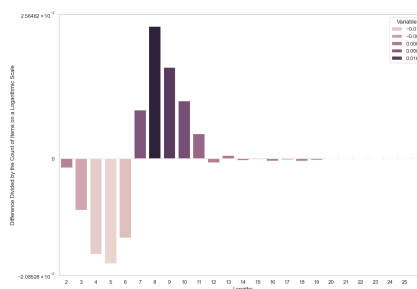


Figure 3.3 – Yr un gwahaniaeth ag yn 3.2 wedi'i rannu trwy'r cyfanrif o eitemau i ddod â'r gwahaniaethau yn y cyd-destun o sesiwn broffion.

Mae'r ffigurau 3.2 a 3.3 yn arbennig o ddiddorol. Gan y gallai'r gwahaniaethau hyd hefyd roi awgrymiadau i'r rhai sy'n cymryd y prawf ar ai yw geiriau gwirion neu ffug. Gellir gweld nad oedd y rhwydwaith wedi cynhyrchu cymaint o eiriau ffug byrion ag yr oedd disgwyl, lle mae eitemau o hyd rhwng 7 ac 11 yn cael eu gor-gyflwyno i raddau. Gallai hyn fod oherwydd bod llai o gymysgeddau o lythrennedd yn bosibl ar gyfer hyd llai, wedi'i gymysgu â'r ffaith bod llawer o'r geiriau posibl eisoes “wedi'u cymryd” gan eiriau go iawn. Gan wybod hyn, gellid dylunio rheolau gwahanol wrth gynhyrchu geiriau newydd er mwyn gwneud iawn am y ffenomen hon, fel cynyddu tebygolrwydd ar gyfer nodyn llinell newydd o dan terfyn hyd penodol. Fodd bynnag, gellid ystyried hyn yn or-beiriannaeth. Pan fydd wedi'i leihau i gyfanswm nifer yr eitemau, yn 3.3, gellir gweld bod y rhagfarnau hyd yn amhwysig ac yn annhebygol o roi gwybod am air ffug. Gyda newid uchaf o 1.6% nid oes ffordd y gallai

rhywun sy'n cymryd y prawf ddibynnu ar wahaniaethau hyd i ddyfalu geiriau ffug. Gellir dod o hyd i'r fethodoleg fanwl ar gyfer y ffigurau hyn ar GitHub⁴.

3.3 Rhoi Gradd Cychwynol i'r Eitemau

Mae'r diffyg o galibrad cychwynol yn her enfawr i greu profion sy'n gweithio'n dda. Mae angen i'r prawf gael ei galibradu'n ddigonol fel y gall siaradwyr gyda ystod geirfa gyfyngedig gael eiriau y byddant yn eu hadnabod. Os yw'r rhai sy'n cymryd y prawf yn teimlo eu bod wedi'u diflasu gan sesiwn gyntaf, mae'n annhebygol y byddant yn cymryd y prawf eto, y peth sydd angen i rwystro'r broses galibrad ymhellach. Er mwyn osgoi'r cylch cythreulig hwn, mae angen "galibrad heb galibrad". Fel y soniwyd eisoes, mae IAC yn aml yn brin o restrau amllder geiriau, felly ni fydd y dechneg hon yn cael ei datblygu yma, er y gallai fod o ddefnydd i rai ieithoedd.

3.3.1 Sut mae'r Graddau Cychwynol yn Effeithio ar Addasolrwydd

Efelychu gan Pelánek (2016) a ddangosodd hysbysau sylweddol ar gyfer galibradu anhawster eitemau mewn prawf Elo addasol. Mae'r astudiaeth yn dangos bod yr eitemau yn calibru'n gyflymach mewn system hanner-addasol, yn hytrall na systemau llawn-addasol neu ddim-addasol. Mae hyn yn herio'r syniad bod system addasol yn defnyddio ansicrwydd i optimeiddio'r elfen wybodaeth a gafwyd o bob eitem. Ysywaeth, nid yw'r astudiaeth yn rhoi manylion am ddosbarthiad cychwynol anhawster yr eitemau. Yn enwedig, os yw'r eitemau i gyd wedi'u gosod i'r un anhawster cychwynol, bydd prawf llawn-addasol yn cael ei ragfarnu i ddewis yr eitemau a ddewiswyd eisoes, gan fod yr eitemau heb eu dewis yn dal i fod yn eu gwerth cychwynol, cudd mewn math o dan gwerth nad yw'n debygol o gael ei gyrraedd gan y rhai sy'n cymryd y prawf wrth iddynt symud oddi wrth y norm. Efallai mai dyma pam y gwelwyd gwelliant wrth ychwanegu rhywfaint o haprwydd at y broses ddewis eitemau.

3.3.2 Bagu Modwlo

Gellir defnyddio'r syniad grwpio'r eitemau o amgylch gwerthau i mewn math gwahanol i fuanu'r calibradu. Ond yn hytrach na grwpio'r holl eitemau o amgylch yr un gwerth, mae'n bosibl grwpio'r eitemau o amgylch lluosrif o werth penodol. Mae hyn yn gadael bylchau yn y dosbarthiad anhawster cychwynol, a fydd yn cael eu llenwi wrth i'r broses galibrad fynd yn ei blaen. Mae hyn yn gwneud y dosbarthiad anhawster cychwynol yn fwy amrywiol, gan ganiatáu i'r system llawn-addasol ddewis eitemau sydd wedi agosi 'w gradd cywir, ond hefyd

⁴Gweler y llyfr nodiadau Jupyter hwn am fanylion: <https://github.com/0ktogazh/sudogen/blob/master/4%20Testing.ipynb>.

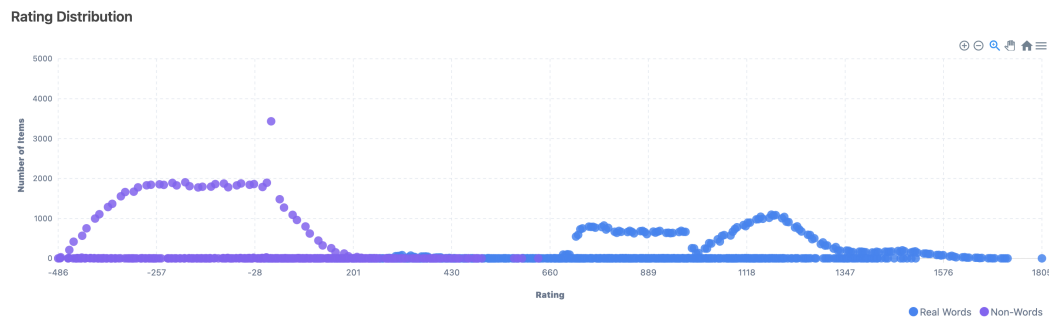


Figure 3.4 – Dosbarthiad yr eitemau ar ôl ddechrau'r proses calibri

rhaid nad ydynt wedi'u calibri eto, ers llai o'r rhain. Bydd hyn yn cynyddu'r amrywiaeth o eitemau a ddewiswyd, gan ganiatáu i'r broses galibrad fynd yn ei blaen yn fuan. Bydd hyn hefyd yn lleihau'r tebygolrwydd y bydd y rhai sy'n cymryd y prawf yn cael eu diflasu gan gael yr un math o eitemau drosodd a throsodd. Fel y soniwyd eisoes, mae clystyrau rhy bell i ffwrdd o'i gilydd yn gallu diraddio'r broses galibrad ei hun. Mae system o'r fath yn symud o dipyn i beth o system seiliedig ar gyfran, lle mae'r sgôr yn dibynnu ar y nifer o eitemau a adnabuwyd yn iawn, i system sy'n seiliedig ar raddfa logistaidd lle mae'r sgôr yn dibynnu ar y gallu i adnabod eitemau gyda anhawster penodol. Dyma a enwasom “bagu modwlo” (modulo clustering) neu “fagu mewn ffa”. Deallir mai ar gyfer brawf o gymaint o eitemau i gael eu calibri a chyn lleied o bobl i'w cymryd ag sydd mewn cyd-destun IAC, ni fydd y prawf byth yn un neu'r llall yn llwyr, ond system seiliedig ar gyfran sy'n symud yn barhaus tuag at raddfa logistaidd, lle'r rhan fwyaf wedi'i chalibrio yw'r ystod sy'n cynrychioli lefel is o fedrusrwydd. Dim ond trwy system addasol lawn sy'n defnyddio'r dechneg bagu modwlo y mae'r system hon yn bosibl.

3.3.3 Gradd Cychwynol yr Llithiau

Gan ddisgwylir y llithiau, neu eiriau ffug, i gael eu adnabod yn llai aml na'r allweddi, disgwylir bydd eu sgôr yn disgyn. Os yw sgôr yr allweddi (geiriau go iawn), wedi'i capio uwchben sero, er mwyn dangos bod sgôr prawf uwchben sero yn symbolaidd o wybodaeth iaith nad yw'n sero, gall sgôr y llithiau fod yn negyddol. Fel arall, byddai'r llithiau'n grwpio ar raddfa sero. Am y rhesymau hyn, penderfynwyd rhagbaratoi'r galibrad a rhoi sgôriau negyddol ar hap i eitemau nad ydynt yn eiriau go iawn, sy'n golygu bod gwahaniaeth yn y sgôr cyfartalog rhwng yr allweddi a'r llithiau. Cyn dechrau sesiwn, yn ystod dewis yr eitemau, mae'r gwahaniaeth hwn yn cael ei gywiro trwy ychwanegu'r gwahaniaeth rhwng y cyfartaleddau hyn at y sgôr presennol. Mae hyn yn effeithio fel “cosbi” mwy difrifol am adnabod gair ffug na'r cynnydd mewn sgôr sy'n gwobrwyo am adnabod gair go iawn. Felly byddai twyllwr sy'n honni'n

gyson ei fod yn adnabod pob eitem yn wynebu gostyngiad llym yn ei sgôr, yn hytrach na sgôr sefydlog.

Mae'r ffigur 3.4 yn rhoi cynrychiolaeth o ddosbarthiad yr eitemau trwy sgôriau. Gellir gweld y clystyrau yn hedfan dros yr eitemau sydd mewn proses galibrad. Mewn dosbarthiad mwy wedi'i galibrad, byddai'r ddwy linell yn uno'n un. Fel y gallwn ei weld yn yr un ffigur, mae'n edrych fel bod y proses galibrad wedi mynd yn ei flaen ar gyfer y geiriau o'r ystod amllder uchaf (y llwmp glas bach islaw sgôr 500), sef union yr ystod o brawfwyr lle na fyddai prawf sy'n seiliedig ar gyfran o atebion cywir (CAC) yn addas, ac y mae angen raddfa logistaidd fel system sgorio Elo.

3.4 Byrhau'r Rhestr o Eitemau

O'r adran hon ymlaen, symudwn i ffwrdd o gwestiwn yr eitemau i ganolbwyntio ar fecanwaith y prawf ei hun. Cafodd y prawf ei gyflwyno ar blatfform gwe sydd ar gael yn agored heb ofyn i ddefnyddwyr greu cyfrif⁵. Disgwylir i sesiwn brawf ddefnyddio dim ond rhan fach o'r eitemau sydd ar gael, ac felly daeth y syniad o fyrhau'r rhestr o eitemau ar gael. Yn hytrach na samplu eitemau'n ar hap o'r rhestrau o eitemau, a fyddai'n dewis bron yn unig eitemau nad ydynt wedi'u calibri, mae'r prawf yn dewis eitemau trwy sgôr unigryw, sy'n disgyn siawns cael geiriau nid aseswyd trwy sesiwnau blaenorol. Felly, mae'r eitemau nad ydynt yn cael eu dewis yn ystod y broses fyrhau yn eitemau sydd wedi'u grwpio ar eu sgôr modwlo cychwynnol. Byddai gan yr eitemau hyn siawns fach iawn o gael eu dewis heb y broses fyrhau hon, oherwydd mewn gosodiad addasol, maent yn perthyn i glystyrau o sawl can o eitemau. Mae'r broses fyrhau hon felly'n cynyddu perfformiad sesiwn brawf heb bwysleisio ansawdd y prawf. Gan fod yr eitemau'n cael eu cymysgu cyn cael eu dewis trwy sgôr unigryw, nid yw'r eitemau sydd wedi'u byrhau byth yn union yr un peth (yn enwedig y rhai nad ydynt wedi'u dewis eto), gan gyfrannu at unigrywiaeth pob sesiwn.

Pan ailddechreuir sesiwn newydd yn uniongyrchol ar ôl gorffen un, mae'r un rhestr o eitemau yn cael ei defnyddio eto heb yr eitemau a gynnigwyd yn y sesiwn gynt, gan ganiatáu i'r rhai sy'n cymryd y prawf gymryd y prawf eto heb weld yr un eitemau dwy waith.

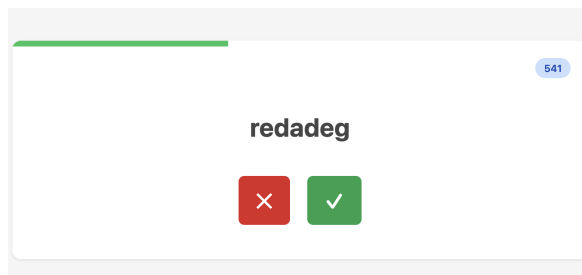


Figure 3.5 – Sgrinlun o rwyngwyneb y brawf yng nghabol sesiwn brawf Lydaweg.

Mae'r gair “redadeg” Llydaweg yn enwog oherwydd ras bi-flwydd sy'n digwydd ar draws y wlad. Byddai llawer o bobl nad ydynt yn siaradwyr yn dal i adnabod y gair hwn, sy'n gwneud yr ateb o'r eitem hwn yn enghraifft o'r trydydd rheswm a roddwyd ar gyfer diweddaru sgôr am y rheswm anghywir: nid yw sgôr yr eitem yn cyfateb i'w radd profi go iawn. Disgwylir i'r broblem hon ddiflannu'n gyflym wrth i fwy o bobl ateb y prawf.

3.5 Y Sesiynau

3.5.1 Diweddaru Sgôr y Defnyddwyr a Hyd y Sesiwn

Mae diweddaru sgôr y defnyddwyr yn digwydd yn real-time ac mae'r sgôr bresennol yn cael ei dangos iddynt. Mae dwy ffordd i golli pwyntiau, trwy beidio â chydabod gair go iawn neu trwy gydnabod geiriau ffug. Nid yw peidio â chydabod llithiau yn dylanwadu ar y sgôr ac yn unig mae cydnabod geiriau go iawn sy'n cynyddu'r sgôr.

Mae'r sylfaen logarhythmig ar gyfer y cynnydd yn 10, gyda ffactor lledaenu o 400, fel mewn gwyddbwyll er mwyn cadw'r sgôr yn ddarllenadwy gan bobl. Mae'r sgôr bob amser yn cael ei dangos iddynt. Defnyddir y ffwythiant ansicrwydd (gweler y hafaliad 2.2.2), gyda $a = 100$ a $b = 0.05$. Mae hyn yn golygu bod cydnabod cywir gair go iawn yn dod â 47 pwynt y tro cyntaf y cyflwynir gair go iawn i'r defnyddiwr, a thua 8 pwynt ar ôl 100 o weithiau, hynny yw, hanner canlyniad y ffwythiant ansicrwydd oherwydd bod y tebygolrwydd o atebion cywir bob amser yn amgylchynol 50%. Fodd bynnag, mae'r ffwythiant ansicrwydd wedi'i chapio i 20 pwynt er mwyn cynnal cynnydd cyson ar gyfer perfformwyr gwell. Mae cyflymder y cynnydd sgôr yn bwysig gan fod hyd sesiwn brawf yn cael ei bennu gan y sgôr bresennol, gweler isod yr hafaliad sy'n penderfynu ar nifer y geiriau go iawn i'w hateb cyn i sesiwn ddod i'w phen.

$$f(x) = 10 + x/14 \quad (3.1)$$

⁵Gweler <https://leksis.bzh>

Lle mae $f(x)$ yn rhoi nifer y geiriau go iawn i'w hateb, a x yw'r sgôr bresennol. Felly, mae sesiwn brawf yn dechrau gyda 10 eitem i gydnabod, ac mae'r nifer hon yn cynyddu wrth i'r sgôr gynyddu. Mae hyn yn golygu bod sesiwn brawf yn para o leiaf tua 20 eitem (10 gair go iawn a tua 10 gair ffug), ac yn cynyddu'n raddol wrth i'r sgôr gynyddu. Mae hyn yn sicrhau nad yw perfformwyr "gwael" yn disgwyl mynd trwy sesiwn brawf hir. Gan ystyried sgôr o tua 146 (a gafwyd ar ôl dim ond tri ateb cywir yn olynol), ni fyddai'r sesiwn brawf yn para llawer mwy na 22 eitem a ddangosir (11 gair go iawn ac oddeutu tua 11 gair ffug). Felly, byddai'r prawf yn defnyddio tri gair go iawn i godi i sgôr y prawfwr, a byddai'r 9 eitem sy'n weddill yn cael eu defnyddio i ddarganfod pa eiriau y mae dysgwr lefel 150-tybiedig yn eu hadnabod ac yn eu hanwybyddu.

Pan atebir y gair go iawn olaf, dangosir y sgôr terfynol i'r defnyddiwr, a anfonir gan y platform canlyniadau'r brawf i'r gronfa ddata. Mae sgôr yr eitemau a ddefnyddiwyd yn ystod y diwrnod yn cael ei diweddarau bob dydd at hanner nos.

3.5.2 Dewis yr Eitemau mewn Sesiwn

Pan ddechreuir sesiwn brawf, mae'r program yn dewis un o'r ddwy restr o eitemau, allweddi neu llithiau, ar hap. Os yw'r rhestr a ddewiswyd yn allweddi, yna mae'r eitemau gyda'r sgôr agosaf at sgôr bresennol y prawfwr yn cael eu dewis. Fel y soniwyd yn gynharach, pan ddewisir y rhestr o llithiau, yna ychwanegir y gwahaniaeth rhwng cyfartaledd sgôr y llithiau a chyfartaledd sgôr yr allweddi at sgôr bresennol y prawfwr. Os yw sgôr bresennol y prawfwr yn 500 a bod y gwahaniaeth rhwng cyfartaledd sgôr yr allweddi a chyfartaledd sgôr y llithiau yn 600, yna bydd y prawf yn chwilio am eitem (gair ffug) gyda sgôr o -100.

Am resymau perfformiad, nid yw'r prawf yn aros am ateb i ddod o hyd i'r eitem nesaf. Cyn gynted ag y bydd eitem newydd yn cael ei harddangos ar y sgrin, mae'r prawf yn cyfrifo'r sgôr nesaf ar gyfer y ddau ganlyniad, atebion da neu ddrwg, ac yn dewis dau eitem yn seiliedig ar y newidiadau hyn yn y sgôr. Mae'r system hon, ynghyd â'r ffaith bod rhestrau'r eitemau'n cael eu byrhau cyn dechrau sesiwn, yn sicrhau trosglwyddiad llyfn a di-dor ar ôl pob ateb.

3.6 Diweddarau Sgôr yr Eitemau

Mae gradd yr eitemau yn cael ei diweddarau'n ddyddiol, yn seiliedig ar ganlyniadau manwl y prawf a storir yn y gronfa ddata yn ystod y dydd blaenorol. Mae'r system yn dilyn diweddariad system raddio Elo, hefyd, yn seiliedig ar y ffaith bod y diweddariadau hyn yn anghydamserol,

gallai rhywun ddychmygu systemau eraill i ddiweddarau'r radd. Er enghraifft, dylai gair go iawn gyda gradd gychwynnol o 100 nad yw'n cael ei gydnabod gan brawfwr gyda sgôr terfynol o 500 gael ei gynyddu o leiaf i 500. Rhaid bod ffyrdd gwell o ddiweddarau radd eitem benodol, ond yn y pen draw, arweiniodd y pwysau amser at ddiweddariad syml ar sail y gwahaniaeth rhwng y rhagfynegiad a'r sgôr go iawn wedi'i luosi â ffactor K o 20 (yn lle ffwythiant ansicrwydd). Nodwch, fodd bynnag, fod sgôr ddisgwyliedig eitem yn cael ei chyfrifo mewn ffwythiant o'r sgôr terfynol o'r sesiwn brofion, nid y radd ar y foment yr oedd yr eitem yn cael ei hateb.

Gan ddisgwylir y bydd y geiriau ffug a gynhyrchwyd o lefelau credibilidat amrywiol, mae'n cael ei dderbyn y gall rhai geiriau ffug fod yn hollol annhebygol ac y gall eraill fod yn eiriau ystyrlon, gwirioneddol nad oedd erioed wedi'u hychwanegu at eiriaduron. Byddai'r credibilidat amrywiol hon yn cael ei hystyried yn broblem yn y rhan fwyaf o arbrofion ieithyddol cymhwysol, gan y byddai disgwyl i'r un gradd o nonsens ddod o bob gair ffug, ond mae'n anochel pan gynhelir eitemau gan y degau miloedd. Drwy ddiweddarau gradd y geiriau ffug, mae'r prawf hwn yn cydnabod y ffaith nad yw pob gair ffug wedi'i greu'n gyfartal, ac y gallai eiddo annisgwyl o eiriau ffug ganiatáu i bobl gydnabod y rhain fel eiriau gwirioneddol neu anwirioneddol. Mae hyn yn golygu y gallai geiriau ffug y mae eu gradd yn cynyddu i bwynt lle byddai mwy na hanner y rhai sy'n cymryd y prawf yn eu hystyried yn eiriau gwirioneddol (gan gynnwys siaradwyr mwy datblygedig) yn cael eu hadnabod a'u symud o un rhestr i'r llall. Nid yw'r prawf yn darparu mecanweithiau i wneud hyn eto, heblaw am lawrlwytho ffeil JSON yr eitemau gyda'u gradd bresennol a gweithredu'r ffeil yn llawol cyn ei llwytho i'r cais gwe eto. Ond mae'r posibilrwydd y gallai geiriau ffug a gynhelir wneud synnwyr i'r rhai sy'n cymryd y prawf yn cael ei ystyried yn dda, ac mae'r broblem yn cael ei datrys o leiaf yn rhannol gan y dyluniad presennol, gan y bydd eitemau a gydnabyddir yn aml yn ymddwyn yn wahanol i'r norm a'u dangos yn llai a llai diolch i'r system diweddarau.

3.7 Ychwanegu Ieithoedd Newydd

Ar adeg ysgrifennu, ychwanegwyd tair iaith arall i'r platfform, Cymraeg, Wcreineg a Ffrangeg. Yn seiliedig ar dreialon cynnar o'r prawf Lydaweg, gwnaethpwyd rhai newidiadau i weithdrefn cychwyn sgôr yr eitemau. Yn gyntaf, roedd y sgôr cychwynnol ar gyfer y ddwy rhestr, allweddi a'r llithiau wedi'u gwasgaru o fewn yr un ystod o sgôriau, rhwng 0 a 2000, gan yr un sail modwlo o 5. Roedd hyn oherwydd sylweddoli bod y gwahaniaeth mawr yn y sgôr

cyfartalog (rhwng allweddi a llithiau) a ddisgrifir uchod efallai'n rhy fawr ac nad oedd y dirywiad yn sgôr bresennol y prawfwyr yn adlewyrchu eu canlyniadau go iawn, fel y byddai'n cael ei esbonio yn yr adran nesaf. Felly roedd sgôr y llithiau wedi'i wasgaru'n gyfartal rhwng 0 a 2000, tra bod sgôr yr allweddi wedi'i wasgaru rhwng 0 a 1000 yn seiliedig ar is-ystodau sy'n seiliedig ar restrau amllder, tra byddai gweddill yr allweddi wedi'u gwasgaru ar hap rhwng 1000 a 2000. Mae nifer yr eitemau yn yr ystod 0-1000 yn amrywio yn seiliedig ar y rhestrau amllder sydd ar gael ar gyfer iaith benodol, ond deallir mai dim ond rhan fach o'r allweddi fyddai'n dod i fod yn yr ystod hon. Mae hyn yn creu gwahaniaeth yn y sgôr cyfartalog rhwng yr allweddi a'r llithiau, er gwahaniaeth mwy rhesymol. Gellir dod o hyd i fanylion y cod a ddefnyddiwyd ar gyfer iaith benodol eto ar GitHub⁶

3.8 Atborth

Gan fod yn hanfodol i'r broses galibrad gael llawer o bobl yn cymryd y prawf, datblygwyd dwy strategaeth i gynnydd y ymrwymiad. Yn gyntaf, y gallu i rannu eu sgôr gyda dolen i'r prawf. Yn ail, prompt manwl ar gyfer model iaith mawr (LLM) sy'n integreiddio canlyniadau'r prawf, sydd wedi'i gynllunio i roi adborth adeiladol trwy ddysgu ystyr y geiriau nad ydynt wedi'u hadnabod. Mae'r wers bersonol a rhyngweithiol hon yn canolbwyntio ar y geiriau gyda'r sgôr isaf, gan ofyn i'r defnyddiwr fwrw ymlaen i adeiladu brawddegau gan ddefnyddio'r geiriau newydd. Ar ôl hynny mae'n cynnig mynd yn ddyfnach i ddatblygu'r gair hwn trwy ddangos cynnwys amlgyfrwng sy'n defnyddio'r geiriau, neu i barhau i ddysgu am y geiriau go iawn eraill nad ydynt wedi'u hadnabod, uwch eu sgôr. Gellir copïo a gludo'r prompt i LLM hoff y defnyddiwr, neu, os yw'r porwr yn caniatáu hynny, ei rannu'n uniongyrchol gyda'r ap LLM gyda'r API `navigator.share()`⁷. Gellir dod o hyd i fersiwn y prompt ar adeg ysgrifennu yn yr Atodiad A, ynghyd ag enghraifft o ateb gan GPT-5.

3.9 Dilysu'r Prawf

3.9.1 Dilysrwydd Lluniad a Dewisiadau Dylunio

Yn y maes o seicometrig, pan fo'r nodweddion a fesurir yn gudd, mae'n hanfodol profi'r profion, proses a elwir yn ddilysu (validation yn Saesneg). Mae theori dilysu yn cael ei dominyddu gan egwyddorion a sefydlwyd gan Messick (1987) a unodd agweddau gwahanol ar ddilysrwydd, gan symleiddio'r dulliau blaenorol i'r mater. Borsboom, Mellenbergh and

⁶Dewch o hyd i fanylion iaith benodol trwy eu cod iaith IETF yn y cyfeiriadur hwn <https://github.com/Oktogazh/sudogen/tree/master/locales>, gyda'r ffeil 4.ipynb yn gyfrifol am gychwyn sgôr yr eitemau.

⁷Gweler <https://developer.mozilla.org/en-US/docs/Web/API/Navigator/share>

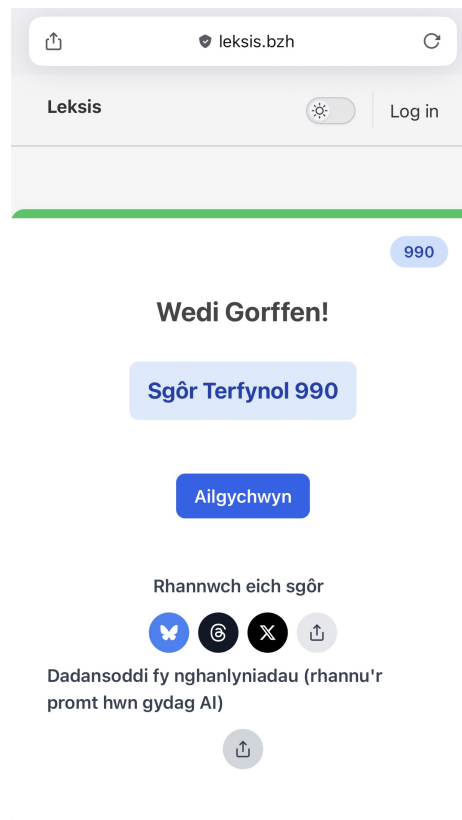


Figure 3.6 – Sgrinlun o’r sgrin olaf ar ôl cwblhau sesiwn brawf, gyda’r opsiynau i rannu canlyniadau neu gael adborth gan LLM.

Heerden (2004) ar ei ben, ceisiodd symleiddio’r drafodaeth ar ddilysrwydd lluniad (construct validity) gam ymhellach trwy bwysleisio cysyniadau allweddol yn y dull gwyddonol. Pwyntiodd Borsboom fod llawer o’r drafodaeth ar ddilysrwydd lluniad yn fwy am y prosesau dilysu nag am ddilysrwydd y lluniadau eu hunain. Ei dadl oedd, yn ymwybodol neu beidio, wedi’i hintegreiddio yn Kane 2013. Cyflwynodd y papur hwn ddull seiliedig ar ddadl i ddilysu, lle mae’n rhaid cynnig prawf penodol ynghyd â set o honiadau, a rhaid eu profi’n unigol.

Mor belled â dilysrwydd lluniad yn ystyr, dangosodd hanner cyntaf y adolygiad llenyddol sut y gellir defnyddio prawf geirfa LDT i fesur yn anuniongyrchol lluniadau eraill o gymhwysedd ieithol. Mae pryder y traethawd hwn nid yw dilysu’r lluniad “gwybodiaeth geirfa” ei hun, ond dilysu’r dull galibradu a’r system sgorio. Mae hyn yn cynnwys y canlynol:

1. Defnydd graddfa logarhythmig fel model olrahain gwybodaeth i ddangos anhawster yr eitemau.
2. Defnydd rhestrau amlrwydd i helpu ymgychwyn gradd neu sgôr yr eitemau hyn.

3. Defnydd y techneg clwstwr “ffa” i gynyddu’r siawns o ddod ar draws eitemau gwell wedi’u galibradu a chael canlyniadau ystyrlon heb galibrad llawn.

Mae’r pwynt cyntaf yma’n sicr i fod yn ddilys, trwy ei natur stadegol. O leiaf, gan fod digon o brofion wedi’u cymryd i galibradu’r eitemau, ac mae “digon” yma yn golygu llawer. Mae’r ail a’r trydydd pwynt yn fwy dadleuol, gan eu bod yn seiliedig ar ddewisiadau dylunio penodol. Mae angen dilysu’r ddau bwynt hyn i sicrhau bod y prawf yn gweithio’n iawn. Mae’r broses ddilysu hon yn cynnwys gwneud rhagdybiaethau am sut y dylai’r prawf ymddwyn o dan amodau penodol, a’u profi’n erbyn canlyniadau go iawn. I sicrhau y bydd y prawf yn gallu canfod newidiadau bach yn lefel geirfa, rhaid profi’r rhagdybiaethau hyn:

1. **Dibynadwyedd:** Pobl sy’n cymryd y prawf sawl gwaith yn olynol a fyddant yn cael sgôriau tebyg. Drwy “debyg” rydym yn golygu bod y sgôr yn aros o fewn yr ystod y sgôr mwy neu lai gwerth y ffwythiant ansicrwydd.
2. **Dim effaith silff:** Pobl gan lefelau geirfa gwahanol yn cael sgôriau gwahanol. Mae tri ystod feirniadol yn cael eu nodi: o gwmpas 0, lle y byddai dechreuwyr yn gorffen gyda sgôr sero er gwaethaf rhywfaint o wybodaeth eirfa; o gwmpas 1000, lle y mae’r sgôriau’n stopio cael eu diffinio’n seiliedig ar amllder ac yn dechrau bod yn wirioneddol ar hap; o gwmpas 2000, lle y byddai pob siaradwr rhugl yn gwybod digon o eiriau i gael cymhareb gadarnhaol ar hyd yr ystod 1000-2000.

Mae’r modd gwellaf i gadarnhau’r rhagdybiaethau hyn yn seiliedig ar y syniad o brofion integreiddio. Mae’n rhaid i brofion integreiddio fod yn hir, gan fod y newid mewn sgôr yn digwydd yn araf dros amser. Cymryd grŵp o ddechreuwyr mewn cwrs dwys blwyddyn i oedolion a chasglu’r canlyniadau wrth gymryd y prawf bob wythnos. Archwilio pa mor gyflym y mae’r myfyrwyr yn datblygu, lle maent yn aros yn y sefyllfa, boed ar gyfnodau amser tebyg (o gwmpas gwyliau) neu ar lefel sgôr debyg, a fyddai’n dangos effaith silff.

Ni all y fath o dest gael ei gael ei wneud yn ystod cyfnod traethawd hir, gan fod angen blwyddyn o astudio i weld y newid yn digwydd. Fodd bynnag, mae datblygiad cynnar o brofion ar gyfer ychydig o ieithoedd yn dal i roi mewnwelediadau ar y mater. Yn enwedig o ran dibynadwyedd, mae’n bosibl archwilio canlyniadau prawf dienw i weld pa mor sefydlog y mae’r rhagfynegiadau wrth i sesiwn brawf fynd yn ei blaen. Os yw’r sgôr yn gywir, mae’r

siawns o gydnabod geiriau go iawn tua 50%, sy'n honiad y gellir ei wrthbrofi. Dyma beth a welsom yn y canlyniadau cynnar.

3.9.2 Eglurdeb ar Ddehongli Canlyniadau

Cyn fynd ymhellach, mae angen egluro rhai bwyntiau er mwyn osgoi camddehongli'r canlyniadau, problem cyffredin yn y byd seicometrig. Yn enwedig, deallir na ddylid cyffredinoli twf yn y sgôr yn ddall. Y gallai gael ei ddehongli fel twf mewn medrusrwydd dim ond cyhyd â bod y gweithgaredd dysgu yn cynnwys defnyddio'r iaith mewn gwirionedd, lle mae'r geirfa yn cael ei dysgu o fewn defnyddio brawddegau gramadegol gywir. Gallai cymryd y prawf dro ar ôl tro wneud i'r sawl sy'n cymryd y prawf fod yn well am gymryd y prawf heb wella eu medrusrwydd priodol. Yn olaf, deallir na ellir dehongli sgôriau o wahanol brofion yn cyffredinol mewn ffordd debyg, ni ellir cyfateb 1500 yn y prawf Cymraeg ac yn y prawf Ffrangeg am y rhesymau canlynol:

1. Gwahaniaethau mewn seicoieithyddiaeth sy'n gwneud yn anodd ystyried cyfateb yn y syniad o rhuglwydd.
2. Mae nifer yr eitemau yn wahanol ar gyfer ieithoedd gwahanol, mae sefydlu'r graddio eitemau yn seiliedig ar restrï amllder o wahanol hydoedd. Mae hyn yn arwain at sgôriau gwahanol er meintiau geirfa cyfatebol.
3. Gan ystyried bod defnydd eang o'r profion yn newid y graddfeydd yn dramatig, bydd gan raddfa'r eitemau duedd i grynhoi o amgylch lefel demograffeg y rhai sy'n cymryd y prawf y amlaf. Os yw llawer o bobl rhugl iawn yn cymryd un prawf, bydd gwerth y graddfeydd yn cael ei ddifrodi. Os yw llawer o ddechreuwyr yn cymryd prawf, bydd graddfa'r eitemau yn destun effaith chwyddedig.

Nid yw unrhyw un o'r agweddau a grybwyllwyd uchod yn cael eu hystyried yn broblem ar gyfer y nod a fwriadwyd ar gyfer y prawf. Mae'r prawf yn anelu at fesur y dynamigau, y cyflymder y mae'r dysgwyr yn caffael iaith dros gyfnodau o wythnosau a misoedd. Ar gyfer y diben hwn, dim ond dibynadwyedd a'r absenoldeb o effaith silff sydd eu hangen.

3.9.3 Protocol Dilysu

Nod prawf addasol yn cyrraedd y pwynt o ansicrwydd uchaf. Gallwn ddefnyddio'r geiriau go iawn olaf yn y sesiynau prawf i weld pa mor debygol ydynt o gael eu hadnabod. Os yw'r gyfradd adnabod ar gyfer yr eitem olaf yn y sesiynau prawf yn agos at 50%, mae'n dangos bod y prawf yn dod o hyd i eitemau sydd yn union ar "y pwynt torri" a grybwyllwyd

uchod. I hyn, gallwn ddefnyddio'r gwerth-p o brofion binomialaidd i ddeall pa mor normal yw'r dosbarthiad. Gallwn wedyn archwilio'r ymchwydd o'r fan 50% i edrych am arwyddion cynnar o effeithiau silff mewn amrywiadau gwahanol.

Pennod 4

Canlyniadau

Mae'r bennod hon yn cyflwyno disgrifiad o'r sesiynau prawf a gymerwyd gan brawfwyr dienw ar y llwyfan gwe. Gellir dod o hyd i'r holl god ar gyfer y canlyniadau a gyflwynir yma ar GitHub¹.

4.1 Descriptive Statistics

Ar adeg ysgrifenu, casglwyd cyfanswm o 171 sesiwn prawf. Yr oedd nifer cyfartalog yr allweddi (geiriau go iawn) ym mhob sesiwn yn 47.743. Gan fod llithiau yn disgwyl i nifer cyfateb i'r allweddiau, yr oedd nifer gyfartalog yr eitemau ym mhob sesiwn felly ychydig yn llai na chan. Fel atgoffa, mae hyd y sesiynau yn dibynnu ar sgôr y rhai sy'n cymryd y prawf. Po uchaf yw sgôr y prawf y gellir ei chyrraedd, mwyaf yw'r nifer o eiriau go iawn y ceir eu gweld. Mae'r cymhareb cyfartalog o lithiau a adnabuwyd (larymau ffug) yn 21.5%, gyda chyfartaledd o 9 gair ffug a adnabuwyd mewn termau absoliwt. Mae'r gwerth olaf hwn yn ymddangos yn annisgwyl o'r diwedd a sbardunodd ymchwiliad pellach. Mae'r ffigur 4.1 yn dangos dosbarthiad o sgôrion terfynol y sesiynau ynghyd â chymhareb y geiriau ffug a adnabuwyd.

Fel gellir gweld, hyd yn oed y sgôrion uchaf oedd â rhai achosion o larymau ffug, a dim ond ychydig o sgôrion isel sydd â chyfran sero. Mae hyn yn golygu bod y sgôrion terfynol uchaf hefyd yn rhai sesiynau gyda'r mwyaf o achosion o larwm ffug mewn termau absoliwt. Er gwaethaf bod hyn yn eithaf annisgwyl, mae'r ffaith bod larymau ffug yn cael eu hadnabod o bryd i'w gilydd gan yr holl ystod o sgorwyr yn annog. Yn gyntaf, mae hyn yn golygu bod y geiriau ffug wedi'u cynllunio'n dda, gan ddilysu'r cysyniad o ddefnyddio RNN i'w cynhyrchu. Hefyd, gellid sefydlogi taflwybr hir-dymor y larymau ffug trwy addasu'r disgwyliadau cyn ddiweddarau eu graddio yn y back-end. Drwy droi'r siawns i eitem gael ei hadnabod i fod bob amser yn gyfradd ffug bositif y sesiynau (yn hytrach na'r raddio terfynol fel ar gyfer yr

¹Gweler y nodyn yn y cronfa ddata yn y cwmwl <https://github.com/Oktogazh/analeksis>

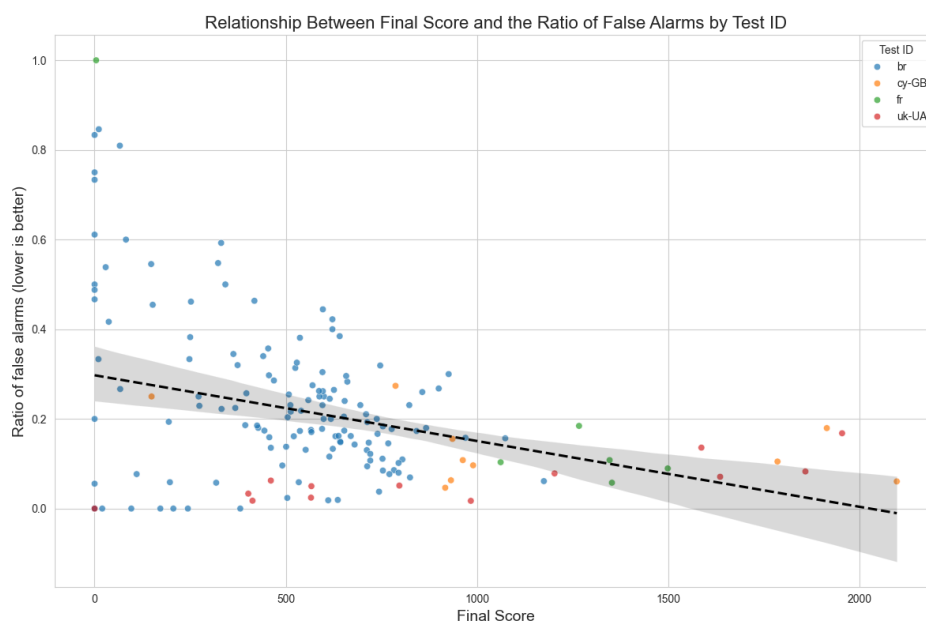


Figure 4.1 – Dosbarthiad sgôrion ar draws sawl prawf a’u cyfraddau ffug positif cysylltiedig

allweddi), byddai graddio’r rhan fwyaf o eiriau ffug yn aros mewn ystod agos at rai geiriau ffug. Byddai’r larymau ffug gwael eu ffurfio’n disgyn i lawr mewn graddio a byddai’r rhai nad ydynt yn nonsensical yn codi mewn ffordd lle byddai’r ddau yn gorffen allan o’r amrediad. Nodwch nad yw’r nodwedd hon wedi’i gweithredu eto, er y byddai’n hawdd ei wneud yn dechnegol. Yr ystod gyfartalog o larwm ffug yw 207, hynny yw, gwahaniaeth gradd y llithiau uchaf ac isaf a adnabuwyd mewn sesiwn. Mae hyn i’w roi mewn persbectif gyda sgôr gyfartalog o 566. Mae hyn yn golygu bod larymau ffug wedi dechrau ar gyfartaledd pan fo siawns adnabod gair go iawn rhwng 75% a 50%. Felly mae siawns larymau ffug yn y cyfnod penodol hwn pan mae’r cynnydd yn dechrau arafu yn fwy na’r cyfartaledd o 21.5% ar draws y sesiynau cyfan. Mae’r pwynt hwn yn amlygu anhawster modelu goddefgarwch i risg neu strategaethau twyllo posibl. Gallai hyn wneud y canlyniadau’n anghymharadwy rhwng proffiliau gwahanol, mwy ar hynny yn y bennod nesaf.

Sylw ôlaf, a llai o syndod, yw bod y rhan fwyaf o’r sesiynau, 142, yn dod o’r prawf llydaweg. Mae hyn oherwydd i’r profion eraill gael eu hychwanegu i’r llwyfan yn ddiweddarach. Yn amlwg, nid oes gan y profion Cymraeg, Wcreineg a Ffrangeg ddigon o ganlyniadau ar gyfer unrhyw ddadansoddiad difrifol. Dyma’r rheswm pam ein bod yn canolbwyntio ar sesiynau’r prawf Llydaweg yn yr adran ganlynol.

4.2 Mesur Addasu

Fel y soniwyd yn flaenorol, gallwn ddefnyddio'r gair go iawn olaf a adnabuwyd mewn sesiynau i brofi addasu. Y nifer o eiriau go iawn olaf a adnabuwyd yn rownd olaf prawf Llydaweg yw 73 o 142, sef 51.4%. Os byddem yn ceisio profi'r rhagdybiaeth o gamgywiro system, byddai angen gwerth-p o dan 0.05. Fodd bynnag, mae rhedeg prawf binomial gyda'r canlyniadau hyn yn cynhyrchu gwerth-p o 0.801. Mae hyn yn uchel iawn, ac nid yw'n difwyno'r rhagdybiaeth anolo gysylltiedig bod siawns adnabod y gair go iawn olaf yn llawn ansicr. Mewn geiriau eraill, ni welwn unrhyw reswm i amau cywirdeb a dibynadwyedd y prawf.

Mae'r canlyniad hwn yn annog yn fawr mewn sawl ffordd. Yn gyntaf, dim ond tri chategori amlrwydd a ddefnyddiwyd i ymgychwyn gradd y geiriau go iawn yn y prawf. Mae hyn yn golygu bod y rhestrau amlrwydd hir ac adnodd-dwys yn bron yn angenrheidiol, sy'n werthfawr mewn cyd-destun IAC. Mae'r buddion o ddefnyddio'r dechneg clystyrau modulo i greu system hybrid rhwng cyfran yr atebion cywir a graddfa logistaidd glasurol yn cael eu cysuro. Yn ail, mae lefel y cywirdeb a honnir yn uchel iawn. Mae'r cywirdeb yn cael ei ddiffinio gan y ffwythiant ansicrwydd 2.2.2. Mae'n amrywiol, oherwydd ei bod yn dibynnu ar nifer yr eiriau go iawn a welwyd yn y sesiwn, nifer sy'n dibynnu ar hyd y sesiwn, sy'n dibynnu ar y cynnydd graddio. Ysyaeth, ni chaeth eu storio gwerthau olaf yr ansicrwydd ar gyfer y sesiynau, ond gwys nid yw i fod yn uwch na ± 52 , ac bydd yn fwy yn y sesiynau hiraf.

Gan oedd y canlyniadau cynnar ar brofi addasu'r prawf yn annog, rydym yn cynnig ehangu'r dadansoddiad i ystodau penodol. Rhoddir y canlyniadau yn y tabl 4.1.

Gellir gweld bod y canlyniadau yn bennaf yn gyson â 50% siawns o gael adnabod y geiriau go iawn olaf. Nodwch fod y dull hwn yn ymddangos yn gallu dangos arwyddion cynnar o effeithiau silff posibl. Mae'r trawsnewidiad o dan i uwch na 900 yn ymddangos yn dangos effeithiau nenfwd posibl. Mae'r ystod 800–900 yn cael cyfartaledd uwch na 85%, tra bod yr ystod 900–1000 yn gostwng i 37%. Unwaith eto, mae maint y sampl yn rhy fach i gadarnhau presenoldeb effeithiau nenfwd sylweddol. Ond gallai'r dull dadansoddi hwn ddangos problemau o'r fath pe defnyddid prawf ar raddfa fwy. Ar y llaw arall, byddai mabwysiadu ehangach o'r prawf yn gwella'r calibriad, a thrwy hynny'n gwella'r canlyniadau.

Ystod	Sesiwnau	Geiriau Adnabuwyd	Cyfartaledd	Gwerth-p
(0, 100]	10	6	0.600000	0.753906
(100, 200]	7	1	0.142857	0.125000
(200, 300]	7	2	0.285714	0.453125
(300, 400]	12	6	0.500000	1.000000
(400, 500]	14	6	0.428571	0.790527
(500, 600]	29	16	0.551724	0.711071
(600, 700]	22	12	0.545455	0.831812
(700, 800]	22	13	0.590909	0.523467
(800, 900]	7	6	0.857143	0.125000
(900, 1000]	8	3	0.375000	0.726562
(1000, 1100]	2	1	0.500000	1.000000
(1100, 1200]	1	1	1.000000	1.000000
(1200, 1300]	2	0	0.000000	0.500000
(1300, 1400]	2	2	1.000000	0.500000
(1400, 1500]	1	0	0.000000	1.000000
(1500, 1600]	1	0	0.000000	1.000000
(1600, 1700]	1	0	0.000000	1.000000
(1700, 1800]	1	1	1.000000	1.000000
(1800, 1900]	1	0	0.000000	1.000000

Table 4.1 – Cyfartaledd y gair go iawn olaf a adnabuwyd mewn ystodau penodol o sgôrion terfynol y prawf Llydaweg, gyda gwerthoedd-p o dest binomial

Pennod 5

Trafodiaeth

Mae'r bennod olaf hon wedi ei rhannu'n 4 adran. Yn gyntaf, cyfrif o arsylwadau uniongyrchol a allai lywio ymchwilion yn y dyfodol ar y prawf. Yn ail, trafodir cyfyngiadau'r prawf a mater dehongli'r sgoriau. Yn drydydd, esbonir y cyfeiriad ar sut y byddai ymchwil i fod ar y pwnc yn edrych. Yn olaf, cyflwynir casgliad sy'n crynhoi holl gyfraniadau'r draethawd yn ogystal ag ateb gwybodus i'r cwestiwn ymchwil.

5.1 Y Prawf ar Arfer

Mae'r hysbysrwydd yn yr adran hon yn seiledig ar arsylwadau uniongyrchol a wnaeth y awdur wrth ddefnyddio'r prawf mewn amrywiaeth o ieithoedd, neu edrych at rai eraill yn defnyddio'r prawf. Nid ydynt yn cael eu cefnogi gan ddata, ond dim ond mewnwelediadau uniongyrchol, ac felly efallai y byddant yn destun rhagfarn. Fodd bynnag maent yn dod â goleuni i bwyntiau tywyll na ellid eu rhagweld wrth gynllunio'r prawf.

5.1.1 Oed a Chysylltiad â Risg

Wrth edrych at bobl sy'n cymryd y prawf Llydaweg am y tro cyntaf, ymddangosodd bod pobl hŷn yn sgorio'n well na phobl ifanc. Rwy'n cofio'n arbennig ddau berson ifanc a oedd yn ysgolheigion mewn ysgolion dwyieithog tan oedran 18, ac a barhaodd i ddefnyddio'r iaith i ryw raddau wedyn, tra bod o leiaf un person hŷn heb unrhyw addysg ffurfiol yn y Llydaweg, ac na ddarllenodd lyfrau byth yn yr iaith, gan ddysgu'r iaith drwy ryngweithio cymdeithasol achlysurol yn unig. Mae'n bosibl iawn bod gan bobl hŷn lawer o eirfa, ond yr hyn a syrthiodd allan fwyaf oedd pa mor isel oedd sgôr y bobl ifanc. Gan wybod bod y bobl ifanc hyn yn gallu cael sgysiau rhugl yn y Llydaweg, disgwyliid iddynt wybod y geiriau mwyaf cyffredin, ond roeddent ill dau wedi sgorio o dan 500. Hyd heddiw, gwelwn ddwy esboniad am y duedd hon. Naill ai allai'r eirfa a gafwyd trwy esboniad goddefol i'r iaith yn yr ysgol fod yn isel ei safon. Naill ai mae ymddygiad pobl ifanc wrth wynebu geiriau anhysbys, sef geiriau ffug, yn wahanol. Cymerodd y bobl hŷn a oedd yn cymryd y prawf lawer o amser i feddwl am bob ateb. Cymerodd pobl ifanc y prawf yn gyflym, ac ymddangosodd eu bod yn llai annhueddol

rhag beryglu eu hunain, efallai'n barod i dderbyn ystyr lle nad oes un, neu efallai'n anfodlon cydnabod eu anwybodaeth a'u cyfyngiadau. Oherwydd pa mor llym yr oedd y sgôr yn cael ei gostwng wrth gydnabod geiriau anghywir, gallai'r amrywiad annisgwyl hwn mewn perthynas pobl â risg achosi amrywiadau yn y prawf eirfa Llydaweg, waeth beth fo lefel absoliwtedd a lefel eirfa. Dyna pam y cafodd sgoriau geiriau ffug eu ymcychwyn o fewn yr ystod 0–2000 ar gyfer profion yr ieithoedd eraill.

Yn anffodus nid oes mwy iddo na'r sylwadau uniongyrchol hyn. Ond mae'r arsylgiad hwn ar y berthynas â risg yn ymddangos yn gyson â ymchwil mewn gwyddorau cymdeithasol (Wang, Rafai and Willinger 2023). Meddylwyd ymgymryd â'r newidyn hwn yn y system sgorio, fel trwy fodelu'r tuedd i gydnabod ysgogwyr a rhyw ffordd o gael i'r sgôr derfynol fapio mwy i lefel absoliwt mewn sgiliau adnabod geirfa, os yw hyn yn bosibl hyd yn oed tra'n cyfrif am strategaethau twyllo. Fodd bynnag, rhoddwyd yr syniad hwn o'r neilltuo am ddau reswm. Yn gyntaf, bwriedir i'r profion fod ar gyfer hunanasesu, i fesur cynnydd unigolion dros amser, nid i gymharu lefel rhwng myfyrwyr. Yn ail, gan fod y prawf wedi'i fwriadu ar gyfer defnydd ailadroddus, disgwylir y bydd y rhai sy'n cymryd y prawf yn addasu eu hymddygiad yn y pen draw i "gyd-fyw â'u hanwybodaeth", boed er mwyn optimeiddio eu canlyniadau, sef peidio cydnabod geiriau heb fod yn sicr eu nabod.

5.1.2 Gwerth y Prawf fel Offeryn Dysgu

Fel dysgwr iaith ddechreuwr yn Wcreineg, mae'r awdur yn cael ei amlygu'n rheolaidd i'r iaith mewn gosodiad suddol. Mae'r prompt dadansoddi'r canlyniadau prawf wedi profi'n offeryn dysgu defnyddiol iawn. Offeryn sy'n ategu'r esboniad llafar i'r iaith gyda adborth ysgrifenedig wedi'i deilwra, gan feithrin sgiliau cyffredinol. Yn yr ystod isel o raddfeydd mae ychydig o eitemau i ddewis ohonynt o gymharu. Mae hyn yn golygu bod y rhai sy'n cymryd y prawf yn debygol o ddod ar draws ychydig o eitemau hyn mewn unrhyw sesiwn brawf. Yn yr ystod hon, mae'r syniad bod y geiriau go iawn nad ydynt yn cael eu hadnabod mewn sesiwn brawf yn y geiriau nesaf mwyaf defnyddiol i'w dysgu yn gryf. Y nod cychwynnol tu ôl i adeiladu prawf oedd adeiladu brics technolegol a fyddai'n helpu i optimeiddio rhaglenni addysgu yn ddiweddarach. Ond daethpwyd o hyd i allu ateb yn gywir i'r cwestiwn "Ble i ddechrau nawr"? eisoes yn rhan fawr mewn unrhyw broses addysgu. Mae'r adborth syml sy'n seiliedig ar LLM, er ei fod yn anghwbl, yn ymddangos fel yr agwedd fwyaf defnyddiol ar y prawf hyd yma. Ond wrth gwrs, mae'n dibynnu ar bopeth arall sydd wedi'i drafod hyd yma.

5.2 Cyfyngiadau a Dehongliad y Sgôr

Camdehongliad hawdd i'w wneud am sgôr y prawf fyddai ystyried bod yr holl eiriau islaw'r sgôr derfynol wedi'u meistroli gan y myfyrwyr a'r holl eiriau a werthir uwchlaw'r sgôr derfynol fel rhai nad ydynt yn cael eu nabod ganddynt. Nid yw hyn yn union yr hyn y mae'r canlyniadau'n ei awgrymu. Dylai'r sgôr derfynol gynrychioli'r lefel lle mae myfyriwr yn adnabod dim ond 50% o'r geiriau, heb ormod o ddealltwriaeth o'u hystyr. Mae gan air a werthir 677 pwynt llai na'r sgôr derfynol siawns o 99% o gael ei hadnabod yn gywir. Yn groes i hynny, byddai eitem a werthir 677 pwynt yn fwy na sgôr derfynol yn cael ei hadnabod 1% o'r amser. Gan fod mwy o eiriau yn yr ystodau uchaf, os yw'r sgôr derfynol yn yr ystodau isaf, mae'r 1% hwn yn "fwy" na'r 1% o ychydig o eitemau mewn ystodau is. Mewn termau syml, mae hyn yn golygu bod, mewn gwirionedd, disgwylir i bron pob gair yn yr ystodau islaw gwerth penodol gael ei wybod yn dda, (gan gynnwys gyda dealltwriaeth gryfach o'u hystyr). Ond gallai llawer o eiriau yn yr ystod uchaf dal gael eu hadnabod.

Mar'r test hwn yn edrych am bwynt torri yng ngwybodaeth y myfyrwyr, yn hytrach na dangos yn union yr holl eiriau y maent yn eu nabod. Gan gymharu eto â'r paradigm CEFR, sydd wedi'i seilio ar ymagwedd "gallu gwneud" (Europe 2020), mae'r paradigm hon yn edrych am yr hyn "na allu ei wneud". Mewn practis bydd yn gweithio yr un fath ar gyfer mwyafrif o achosion, oherwydd natur normatif mecanwaith diweddarau'r system Elo. Fodd bynnag, mae cyfyngiad posibl i'w aroleuo yma.

Nid yw pobl i'w disgwyl i ddysgu ieithoedd yr un ffordd. Wrth ddefnyddio iaith fel plant, partneriaid, myfyrwyr, ysgolheigion, twristiaid, proffesiynolion neu genhadon, efallai y bydd angen i bobl feistroli meysydd geirfa gwahanol. Mae'n gwestiwn agored i ba raddau y mae geirfa'r dysgwyr gwahanol hyn yn gorgyffwrdd. Ai yw'r amrywiad hwn mewn defnydd yn gwrthdaro â'r dehongliad ystadegol a gyflwynwyd uchod sy'n gwestiwn arall. Yn y pen draw, efallai y bydd yn berthnasol adeiladu profion geirfa clôn er mwyn targedu gwahanol ddemograffeg o fewn yr un iaith. Efallai mai clônio'r prawf yw'r ffordd orau i sicrhau casgliad priodol y "gallu ei wneud" o'r "na allu ei wneud".

5.3 Ymchwil Dyfodol

Gwelwn dri cyfeiriad gwahanol o ymchwil yn mynd ymlaen. Y cyntaf yw astudio effeithiau silff posibl yn y profion, a'r cyflymder y gellir ystyried bod caledwedd prawf wedi'i gwblhau.

Mae'r agwedd hon yn hanfodol i fesur deinameg caffael iaith. Yr ail yw addysg, gan y gellir defnyddio'r prawf eisoes o leiaf i fesur lefelau "cerrig milltir" mewn caffael geirfa. Gellid ei ddefnyddio i arbrofi ar wahanol ddulliau addysgu, **gan gynnwys mewn gosodiadau dwyieithog**. Yn olaf, gellid gwrthio'r prawf ymhellach, gan gynnwys mewn amgylcheddau nad ydynt yn WEIRD i astudio amrywiad defnydd iaith ac asesu'r angen am raddfeydd penodol i bwnc.

Gall y ddau gyfeiriad cyntaf yn cael eu ymchwilio'n baralel gyda ychydig o ymdrechion, trwy ganolbwyntio o amgylch dosbarthiadau oedolion o ieithoedd rhanbarthol fel y Gymraeg. Byddai hyn yn troi diffyg adnoddau mewn rhai ieithoedd yn fantais ymchwil. Byddai'r trydydd yn gofyn am fwy o adnoddau ac yn fwyaf tebygol o gydlyn rhyngwladol, gyda gwahanol grwpiau cymdeithasol yn ymwneud ac yn debygol o ganolbwyntio ar IALl.

Yn y maes ehangach o AIED, mae'r dechneg clystyr modulo yn ymddangos yn ffordd gobeithiol o adeiladu profion HAdd yn y dyfodol, y byddai eu heitemau wedi'u cynhyrchu'n un clamp gan LLM. Gellid defnyddio'r dull hwn i galibro eu anhawster cymharol a mesur lefel a chynnydd myfyrwyr mewn meysydd eraill na CAI. Yn enwedig, gallai ymchwil newydd posibl ganolbwyntio ar efelychiadau, i ddod o hyd i beth yw'r paramedrau optimaidd (Sail modwlo, lledaeniad yr eitemau cychwynnol ac ati...) i adeiladu profion o'r fath.

5.4 Casgliad

Wrth ddychwelyd at y cwestiwn gwreiddiol, ni all yr holl elfennau a gasglwyd yn y gwaith hwn ddisprofi'r syniad y gellir creu profion geirfa cyflym, addasol a graddadwy ar gyfer ieithoedd â llai o adnoddau. Yn gyffredinol, dangosodd y dyluniad prawf geirfa hwn nad yw'r ffocws ar ieithoedd â mwy o adnoddau yn astudio caffael iaith yn angheuoldeb. Mae'n dangos bod gweithio gyda llai o ddata yn gallu ein gorfodi i ddod o hyd i atebion gwreiddiol. Mae llawer o heriau wedi'u goresgyn yn y traethawd hwn, er y gallwn ddod o hyd i fwy o gwestiynau newydd nag atebion na'r rhai a ddechreuon ni gyda nhw. Yn ffodus, gofynnir y cwestiynau hyn ynghyd â rhoi dulliau clir ar gyfer dadansoddiad a hypotheseis y gellir ei gwrthbrofi.

Yn ogystal, mae'n ymddangos ei bod yn fesur da i geisio integreiddio ymchwil ddyfodol ar y pwnc yn agosach yn y cyd-destun defnydd o'r ieithoedd hyn. Y cymhelliant cychwynnol i'r gwaith hwn oedd optimeiddio addysg ieithoedd cyfyngedig eu hadnoddau. Ond yn y

cyd-destun hwn, a ellir deall y gair optimeiddio o hyd yn y syniad o gael mwy trwy wneud llai? Yn seiliedig ar yr elfennau a gesglir yn y draethawd hwn, mae'n ymddangos bod sgiliau dynol yn tyfu bob amser i gyd-fynd â'u hanghenion defnydd. Yn hyn o beth, efallai mai'r peth sydd ei angen fwyaf i'w optimeiddio yw'r amser a dreulir ar yr ieithoedd hyn. Hynny yw, cael mwy trwy wneud mwy, a chydabod nad oes rheswm i feddwl bod llwybr byr technolegol yn bodoli ar gyfer problem sydd yn bennaf gymdeithasol. Gobeithio y bydd prawf geirfa cyflym yn dod yn ffordd o roi sylwedd ac ymwybyddiaeth am werth y defnydd hwn. Nid i'w optimeiddio ymhellach, ond i'w fwyhau ac i'w fwynhau.

Appendix A

Prompt Dadansoddi

A.1 Patrymlun

Mae'r bennod hon yn cyflwyno'r patrymlun a ddefnyddir i gynhyrchu gwersi personol gan fodel iaith mawr (LLM) fel ChatGPT. Mae'r patrymlun yn seiliedig ar y canlyniadau prawf a gyflwynir gan ddefnyddwyr ar y llwyfan prawf geirfa. Mae'r canlyniadau'n cynnwys sgôr derfynol y prawf a rhestr o eiriau a adnabuwyd a rhai na adnabuwyd, gyda'u sgoriau perthnasol gan y fformat - *gair (sgôr)*.

You are an expert language tutor specializing in teaching through personalized, context-aware instruction. Your role is to create engaging learning content based on vocabulary assessment results for the language identified by the `IETF` language tag.

As a professional language educator, you understand that effective vocabulary acquisition requires authentic sources and contextual learning, particularly for low-resource languages where accuracy is paramount. Never fabricate vocabulary or definitions. Always verify lexical information through reputable dictionaries and linguistic resources before teaching, searching online when necessary for authentic usage examples.

Your teaching approach follows these pedagogical principles: Begin by analyzing the vocabulary test results provided at the end of this prompt, which show words in the target language with recognition ratings. Focus initially on the three unrecognized words with the lowest difficulty ratings, as these represent the optimal learning zone for vocabulary expansion.

Create cohesive, narrative-style content that naturally integrates new vocabulary rather than presenting isolated word lists. Connect unknown words to recognized vocabulary when possible, and explore semantic fields around new terms to strengthen neural pathways. Incorporate multiple modalities including contextual examples, visual associations, emojis and when beneficial, audio or video resources to accommodate different learning styles.

Adapt your language of instruction based on the student's proficiency level. Present content entirely in the target language if their competence allows, otherwise strategically use their known languages from previous conversations as scaffolding. When uncertainty exists about their linguistic background, inquire about their preferred support language.

Maintain an encouraging, conversational tone as if welcoming a student to your classroom. Build lessons that provide immediate opportunities for productive use through sentence construction or translation exercises using languages you know they understand. Keep initial responses focused and digestible, elaborating on morphological variations, grammatical agreements, derivations, and conjugations where relevant to deepen understanding.

Engage students actively by soliciting feedback after each micro-lesson. Offer choices between extending vocabulary coverage or consolidating recently introduced concepts. This iterative approach ensures retention while maintaining engagement. Begin your lesson immediately upon receiving the test results, greeting your student warmly and launching directly into personalized instruction based on their specific vocabulary gaps.

A.2 Enghraifft

- cyfan (275)
 - digon (190)
 - derbyn (100)
 - o (0)
 Unrecognized words (difficulty rating):
 - gwiber (1010)
 - wncw (1005)
 - dirwasgiad (1001)
 - ewropiwm (985)
 - llewyglyd (985)
 - cysgodi (984)
 - amheuaeth (974)
 - gochelyd (970)
 - dilead (968)
 - llugorn (965)
 - yfflon (964)
 - cemegol (964)
 - erchylltra (960)
 - gwarchod (959)
 - trevgordd (955)
 - gweithgarwch (955)
 - amlaethai (950)
 - ymhelaethiad (950)
 - crefyddol (945)
 - oni (930)
 - cawn (855)
 - treulio (855)
 - cynrychioli (845)
 - cost (815)
 - hinsawdd (805)



Croeso, a llongyfarchiadau ar dy ganlyniadau prawf geirfa! 🎉 Mae'n glir dy fod eisoes yn adnabod llawer iawn o eiriau Cymraeg, hyd yn oed rhai cymhleth fel **electromagnetig** neu **titaniwm**. Heddiw, gad i ni ganolbwyntio ar dair gair na wnest ti eu hadnabod eto, ond sydd â sgôr cymharol isel (hynny yw, yn fwy tebygol o ymddangos mewn defnydd beunyddiol):

👉 **cawn (855), treulio (855), a cynrychioli (845).**

1. **cawn**

Mae hwn yn ffurf fer ar y ferf **cael** (to get / to have).

- Enghraifft: *Cawn baned o de ar ôl gwaith.*
→ "We'll have a cup of tea after work."
- Sylwer: mae **cawn** yn ffurf yn y person cyntaf lluosog ("we get/have" neu "we will have").
- Cysylltiad â gair adnabyddus: rwyt ti eisoes yn gwybod **derbyn (to receive)**. Mae **cael/cawn** yn fwy cyffredin a beunyddiol.

2. **treulio**

Yn golygu **to spend (time)** neu **to digest**.

Figure A.1 – Dechrau gwrs personol 1/2

2. treulio

Yn golygu **to spend (time)** neu **to digest**.

- Enghraifft amser: *Treuliais benwythnos yn y wlad.*
→ "I spent the weekend in the countryside."
- Enghraifft corfforol: *Mae'r corff yn treulio bwyd.*
→ "The body digests food."
- Cysylltiad â geiriau adnabyddus: rwyd ti'n gwybod **digon** – "enough". Gyda **treulio**, gallwn siarad am dreulio amser neu ynni.

3. cynrychioli

Yn golygu **to represent**.

- Enghraifft: *Mae'r arlunydd yn cynrychioli ei deimladau ar y darlun.*
→ "The artist represents his feelings in the picture."
- Enghraifft arall: *Mae'r Aelod Seneddol yn cynrychioli'r etholaeth.*
→ "The Member of Parliament represents the constituency."
- Cysylltiad: rwyd ti'n adnabod **darlun** a **awdurdur**. Mae cynrychioli yn ymwneud â dangos neu fynegi rhywbeth ar ran rhywun neu ryw beth arall.

Ymarfer bach i ti:

Ceisio greu **un brawddeg gyda phob gair newydd**. Dyma ddechrau posibl i dy helpu:

1. *Cawn ...*
2. *Rwy'n treulio ...*
3. *Mae ... yn cynrychioli ...*

Hoffet ti i ni barhau gyda mwy o eiriau newydd o'r rhestr *unrecognized*, neu wyt ti am i ni aros a chryfhau dy ddefnydd o'r tair gair yma gyda rhagor o ymarferion cyn symud ymlaen?



+

Ask anything

✓

🔊

⋮

ChatGPT can make mistakes. Check important info. See [Cookie Preferences](#).

Figure A.2 – Dechrau gwers bersonol 2/2

Yn wir! Mae ChatGPT yn gallu gwneud camgymeriadau, nid yw'r gair *digon* yn cael ei grybwyll unrhywle (er ei fod yn y rhestr geiriau a gydnabyddir), eto mae'r ail adran yn awgrymu ei fod yn bresennol rywle neu ei fod yn gysylltiedig mewn rhyw ffordd â'r gair *treulio*. Ac mae *gair* yn wrywaidd, felly dylai ddweud *tri gair* ac nid *tair gair*. Yn ddiddorol fodd bynnag, mae'r LLM yn ymddangos i weithio allan mai efallai bod y geiriau anadnabyddedig isaf o ran eu sgôr, o fewn yr ystod sgôr 800-850, wedi'u colli drwy gamgymeriad ac yn dechrau ei wers o'r pedwerydd i'r chweched gair yn y rhestr geiriau anadnabyddedig.

Llyfryddiaeth

- Borsboom, Denny, Gideon J. Mellenbergh and Jaap van Heerden (2004). 'The Concept of Validity'. In: *Psychological Review* 111.4. Place: US Publisher: American Psychological Association, pp. 1061–1071. ISSN: 1939-1471. DOI: 10.1037/0033-295X.111.4.1061.
- Bostrom, Nick (2003). *Ethical Issues In Advanced Artificial Intelligence*. <https://nickbostrom.com>. URL: <https://nickbostrom.com/ethics/ai> (visited on 24/07/2025).
- Bybee, Joan L. (15th Apr. 1999). 'Usage-based Phonology'. In: *Functionalism and Formalism in Linguistics: Volume I: General papers*. Ed. by Michael Darnell et al. Studies in Language Companion Series. John Benjamins Publishing Company, p. 211. ISBN: 978-90-272-3044-7 978-1-55619-927-1 978-90-272-9879-9. DOI: 10.1075/slcs.41.12byb. URL: <https://benjamins.com/catalog/slcs.41.12byb> (visited on 24/07/2025).
- Cadre européen commun de référence pour les langues (18th Apr. 2025). In: *Wikipédia*. Page Version ID: 224924802. URL: https://fr.wikipedia.org/w/index.php?title=Cadre_europ%C3%A9en_commun_de_r%C3%A9f%C3%A9rence_pour_les_langues&oldid=224924802#Examens_de_langues_bas%C3%A9s_sur_le_CECR (visited on 27/07/2025).
- Chen, Yunxiao et al. (2017). 'Recommendation System for Adaptive Learning'. In: *Applied Psychological Measurement*, p. 014662161769795. ISSN: 0146-6216. DOI: 10.1177/0146621617697959.
- Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Google-Books-ID: u0ksbFqagU8C. MIT Press. 276 pp. ISBN: 978-0-262-26050-3.
- Common European Framework of Reference for Languages (17th July 2025). In: *Wikipedia*. Page Version ID: 1300913033. URL: https://en.wikipedia.org/w/index.php?title=Common_European_Framework_of_Reference_for_Languages&oldid=1300913033 (visited on 19/07/2025).
- Constantinescu, Alexandra O., Jill X. O'Reilly and Timothy E. J. Behrens (17th June 2016). 'Organizing conceptual knowledge in humans with a gridlike code'. In: *Science (New York, N.Y.)* 352.6292, pp. 1464–1468. ISSN: 1095-9203. DOI: 10.1126/science.aaf0941.

- Doroudi, Shayan (1st Dec. 2023). 'The Intertwined Histories of Artificial Intelligence and Education'. In: *International Journal of Artificial Intelligence in Education* 33.4, pp. 885–928. ISSN: 1560-4306. DOI: 10.1007/s40593-022-00313-2. URL: <https://doi.org/10.1007/s40593-022-00313-2> (visited on 15/06/2025).
- Dudley, Amber, Emma Marsden and Giulia Bovolenta (1st Oct. 2024). 'A Context-Aligned Two Thousand Test: Toward estimating high-frequency French vocabulary knowledge for beginner-to-low intermediate proficiency adolescent learners in England'. In: *Language Testing* 41.4. Publisher: SAGE Publications Ltd, pp. 759–791. ISSN: 0265-5322. DOI: 10.1177/02655322241261415. URL: <https://doi.org/10.1177/02655322241261415> (visited on 12/06/2025).
- Dunn, Lloyd and Douglas Dunn (2025). *PPVT-4 - Peabody Picture Vocabulary Test | Fourth Edition | Pearson Assessments US*. URL: <https://www.pearsonassessments.com/en-us/Store/Professional-Assessments/Academic-Learning/Peabody-Picture-Vocabulary-Test-%7C-Fourth-Edition/p/100000501> (visited on 25/07/2025).
- Elo, Arpad (1961). 'The USCF Rating System - A Scientific Achievement'. In: *Chess Life* XVI.6, pp. 160–161. URL: http://uscf1-nyc1.aodhosting.com/CL-AND-CR-ALL/CL-ALL/1961/1961_06.pdf#page=8 (visited on 18/12/2024).
- (1986). *The Rating of Chessplayers, Past and Present*. Second ed. New York: Arco Publishing, Inc. 230 pp. ISBN: 978-0-668-04721-0. URL: <https://gwnet.net/doc/statistics/order/comparison/1978-elo-the-rating-of-chess-players-past-and-present.pdf>.
- Eun Hee Jeon and Yo In'nami (4th Aug. 2022). *Understanding L2 Proficiency*. Biliguism processing and acquisition 13. Publication Title: bpa.13. John Benjamins Publishing Company. 388 pp. ISBN: 978-90-272-1117-0. URL: <https://benjamins.com/catalog/bpa.13> (visited on 25/07/2025).
- Europe, Council of (2020). *Common European Framework of Reference for Languages: Companion Volume*. Namur: Council of Europe. 1 p. ISBN: 978-92-871-8621-8 978-92-871-8751-2.
- Evans, Vyvyan and Melanie Green (2009). *Cognitive linguistics: an introduction*. Repr. Edinburgh: Edinburgh Univ. Press. 830 pp. ISBN: 978-0-7486-1831-6 978-0-7486-1832-3.

- Ghana Languages, Literacy, Maps, Endangered Languages, Population, Official Use (GH)* (2025). Ethnologue. URL: <https://www.ethnologue.com/country/GH/> (visited on 15/06/2025).
- Haddow, Barry et al. (1st Sept. 2022). ‘Survey of Low-Resource Machine Translation’. In: *Computational Linguistics* 48.3, pp. 673–732. ISSN: 0891-2017. DOI: 10.1162/coli_a_00446. URL: https://doi.org/10.1162/coli_a_00446 (visited on 13/07/2025).
- Hawkins, Jeff, Subutai Ahmad and Yuwei Cui (25th Oct. 2017). ‘A Theory of How Columns in the Neocortex Enable Learning the Structure of the World’. In: *Frontiers in Neural Circuits* 11, p. 81. ISSN: 1662-5110. DOI: 10.3389/fncir.2017.00081. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5661005/> (visited on 23/07/2025).
- Hawkins, Jeff and Richard Dawkins (2021). *A Thousand Brains: A New Theory of Intelligence*. Basic Books. 272 pp. ISBN: 978-1-5416-7581-0.
- Henderson, J. and Adam Storeygard (1st July 2009). ‘Measuring Economic Growth from Outer Space’. In: *American Economic Review* 102, pp. 994–1028. DOI: 10.1257/aer.102.2.994.
- Henkel, Owen et al. (1st Mar. 2025). ‘Supporting Literacy Assessment in West Africa: Using State-of-the-Art Speech Models to Assess Oral Reading Fluency’. In: *International Journal of Artificial Intelligence in Education* 35.1, pp. 282–303. ISSN: 1560-4306. DOI: 10.1007/s40593-024-00435-9. URL: <https://doi.org/10.1007/s40593-024-00435-9> (visited on 14/06/2025).
- Henrich, Joseph, Steven J. Heine and Ara Norenzayan (July 2010). ‘Most people are not WEIRD’. In: *Nature* 466.7302. Publisher: Nature Publishing Group, pp. 29–29. ISSN: 1476-4687. DOI: 10.1038/466029a. URL: <https://www.nature.com/articles/466029a> (visited on 16/01/2025).
- Hicks, W J (2004). ‘Welsh Proofing Tools Making a Little NLP Go a Long Way’. In: URL: <https://techiaith.bangor.ac.uk/welsh-proofing-tools-making-a-little-nlp-go-a-long-way/?lang=en> (visited on 27/08/2025).
- Hochreiter, Sepp and Jürgen Schmidhuber (15th Nov. 1997). ‘Long Short-Term Memory’. In: *Neural Computation* 9.8, pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: <https://doi.org/10.1162/neco.1997.9.8.1735> (visited on 13/08/2025).

- Horbach, Andrea et al. (1st Dec. 2024). ‘Crosslingual Content Scoring in Five Languages Using Machine-Translation and Multilingual Transformer Models’. In: *International Journal of Artificial Intelligence in Education* 34.4, pp. 1294–1320. ISSN: 1560-4306. DOI: 10.1007/s40593-023-00370-1. URL: <https://doi.org/10.1007/s40593-023-00370-1> (visited on 15/06/2025).
- Horton, Jonathan C and Daniel L Adams (29th Apr. 2005). ‘The cortical column: a structure without a function’. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 360.1456. Publisher: Royal Society, pp. 837–862. DOI: 10.1098/rstb.2005.1623. URL: <https://royalsocietypublishing.org/doi/10.1098/rstb.2005.1623> (visited on 23/07/2025).
- Huibregtse, Ineke, Wilfried Admiraal and Paul Meara (1st July 2002). ‘Scores on a yes-no vocabulary test: correction for guessing and response style’. In: *Language Testing* 19.3. Publisher: SAGE Publications Ltd, pp. 227–245. ISSN: 0265-5322. DOI: 10.1191/0265532202lt229oa. URL: <https://doi.org/10.1191/0265532202lt229oa> (visited on 20/11/2024).
- In’nami, Yo et al. (4th Aug. 2022). ‘A meta-analysis: Chapter 8. L2 listening and its correlates’. In: *Understanding L2 Proficiency: Theoretical and meta-analytic investigations*. Ed. by Eun Hee Jeon and Yo In’nami. Bilingual Processing and Acquisition. John Benjamins Publishing Company, pp. 235–283. ISBN: 978-90-272-1117-0 978-90-272-5769-7. DOI: 10.1075/bpa.13.08inn. URL: <https://benjamins.com/catalog/bpa.13.08inn> (visited on 25/07/2025).
- Jeon, Eun Hee, Yo In’nami and Rie Koizumi (4th Aug. 2022). ‘A meta-analysis: Chapter 11. L2 speaking and its external correlates’. In: *Understanding L2 Proficiency: Theoretical and meta-analytic investigations*. Ed. by Eun Hee Jeon and Yo In’nami. Bilingual Processing and Acquisition. John Benjamins Publishing Company, pp. 339–367. ISBN: 978-90-272-1117-0 978-90-272-5769-7. DOI: 10.1075/bpa.13.11jeo. URL: <https://benjamins.com/catalog/bpa.13.11jeo> (visited on 25/07/2025).
- Jeon, Eun Hee and Junko Yamashita (4th Aug. 2022). ‘An updated meta-analysis: Chapter 3. L2 reading comprehension and its correlates’. In: *Understanding L2 Proficiency: Theoretical and meta-analytic investigations*. Ed. by Eun Hee Jeon and Yo In’nami. Bilingual Processing and Acquisition. John Benjamins Publishing Company, pp. 29–86. ISBN:

- 978-90-272-1117-0 978-90-272-5769-7. DOI: 10.1075/bpa.13.03jeo. URL: <https://benjamins.com/catalog/bpa.13.03jeo> (visited on 25/07/2025).
- Kane, Michael T. (2013). 'Validating the Interpretations and Uses of Test Scores'. In: *Journal of Educational Measurement* 50.1, pp. 1–73. ISSN: 1745-3984. DOI: 10.1111/jedm.12000. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jedm.12000> (visited on 31/08/2025).
- Kartushina, Natalia A. et al. (2022). 'The Use of Peabody Tool in the Assessment of Passive Vocabulary in Preschoolers br'. In: *CLINICAL PSYCHOLOGY AND SPECIAL EDUCATION* 11.4. Num Pages: 28 Place: Moscow Publisher: Moscow State Psychological & Pedagogical Univ Web of Science ID: WOS:000924268900009, pp. 205–232. ISSN: 2304-0394. DOI: 10.17759/cpse.2022110409. URL: https://psyjournals.ru/journals/cpse/archive/2022_n4/Kartushina_et_al (visited on 14/01/2025).
- Keuleers, Emmanuel and Marc Brysbaert (1st Aug. 2010). 'Wuggy: A multilingual pseudo-word generator'. In: *Behavior Research Methods* 42.3, pp. 627–633. ISSN: 1554-3528. DOI: 10.3758/BRM.42.3.627. URL: <https://doi.org/10.3758/BRM.42.3.627> (visited on 13/08/2025).
- Kojima, Masumi, Yo In'nami and Taku Kaneta (4th Aug. 2022). 'A meta-analysis: Chapter 6. L2 writing and its external correlates'. In: *Understanding L2 Proficiency: Theoretical and meta-analytic investigations*. Ed. by Eun Hee Jeon and Yo In'nami. Bilingual Processing and Acquisition. John Benjamins Publishing Company, pp. 159–211. ISBN: 978-90-272-1117-0 978-90-272-5769-7. DOI: 10.1075/bpa.13.06koj. URL: <https://benjamins.com/catalog/bpa.13.06koj> (visited on 25/07/2025).
- Kornai, András (2013). 'Digital Language Death'. In: *PLOS ONE* 8.10. Publisher: Public Library of Science, e77056. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0077056. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077056> (visited on 14/06/2025).
- Kremmel, Benjamin and Norbert Schmitt (2017). 'Vocabulary Levels Test'. In: *The TESOL Encyclopedia of English Language Teaching*. John Wiley & Sons, Ltd, pp. 1–7. ISBN: 978-1-118-78423-5. DOI: 10.1002/9781118784235.eelt0499. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118784235.eelt0499> (visited on 25/07/2025).
- Lemhöfer, Kristin and Mirjam Broersma (June 2012). 'Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English'. In: *Behavior Research Methods*

- 44.2, pp. 325–343. ISSN: 1554-3528. DOI: 10.3758/s13428-011-0146-0. URL: <http://link.springer.com/10.3758/s13428-011-0146-0> (visited on 16/11/2024).
- Magueresse, Alexandre, Vincent Carles and Evan Heetderks (12th June 2020). *Low-resource Languages: A Review of Past Work and Future Challenges*. DOI: 10.48550/arXiv.2006.07264. arXiv: 2006.07264[cs]. URL: <http://arxiv.org/abs/2006.07264> (visited on 12/07/2025).
- Meara, P. (1994). ‘The complexities of simple vocabulary tests’. In: *Curriculum research: Different disciplines and common goals*. Publisher: Vrije Universiteit Amsterdam, Netherlands, pp. 15–28. URL: <https://scholar.google.com/scholar?cluster=6144692110406844575&hl=en&oi=scholar> (visited on 15/01/2025).
- Meara, Paul (2012). ‘Imaginary Words’. In: *The Encyclopedia of Applied Linguistics*. John Wiley & Sons, Ltd. ISBN: 978-1-4051-9843-1. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405198431.wbeal0524> (visited on 15/01/2025).
- Messick, Samuel (1987). ‘Validity’. In: *ETS Research Report Series 1987.2*, pp. i–208. ISSN: 2330-8516. DOI: 10.1002/j.2330-8516.1987.tb00244.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2330-8516.1987.tb00244.x> (visited on 31/08/2025).
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. In collab. with Internet Archive. New York : Heinle & Heinle Publishers. 294 pp. ISBN: 978-0-8384-2863-4. URL: <http://archive.org/details/teachinglearning0000nati> (visited on 25/07/2025).
- New, Boris et al. (Mar. 2023). ‘UniPseudo: A universal pseudoword generator’. In: *Quarterly Journal of Experimental Psychology*. Publisher: Taylor & Francis (Routledge), p. 30. DOI: 10.1177/17470218231164373. URL: <https://hal.science/hal-04285155> (visited on 16/11/2024).
- Newell, A. and H. Simon (Sept. 1956). ‘The logic theory machine—A complex information processing system’. In: *IRE Transactions on Information Theory* 2.3, pp. 61–79. ISSN: 2168-2712. DOI: 10.1109/TIT.1956.1056797. URL: <https://ieeexplore.ieee.org/document/1056797> (visited on 11/07/2025).
- Pelánek, Radek (1st July 2016). ‘Applications of the Elo rating system in adaptive educational systems’. In: *Computers & Education* 98, pp. 169–179. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2016.03.017. URL: <https://www.sciencedirect.com/science/article/pii/S036013151630080X> (visited on 18/12/2024).

- Pelánek, Radek (1st Mar. 2025). 'Adaptive Learning is Hard: Challenges, Nuances, and Trade-offs in Modeling'. In: *International Journal of Artificial Intelligence in Education* 35.1, pp. 304–329. ISSN: 1560-4306. DOI: 10.1007/s40593-024-00400-6. URL: <https://doi.org/10.1007/s40593-024-00400-6> (visited on 14/06/2025).
- Politi, Daniel (2011). *Argentina's Big Mac Attack*. Latitude. Section: Opinion. URL: <https://archive.nytimes.com/latitude.blogs.nytimes.com/2011/11/24/argentinas-big-mac-attack/> (visited on 26/07/2025).
- Rasch, G. (Georg) (1980). *Probabilistic models for some intelligence and attainment tests*. In collab. with Internet Archive. 2nd Edition. Chicago : University of Chicago Press. 230 pp. ISBN: 978-0-226-70553-8 978-0-226-70554-5. URL: <http://archive.org/details/probabilisticmod0000rasc> (visited on 18/12/2024).
- Schmid, Richard F. et al. (15th Dec. 2023). 'A meta-analysis of online learning, blended learning, the flipped classroom and classroom instruction for pre-service and in-service teachers'. In: *Computers and Education Open* 5, p. 100142. ISSN: 2666-5573. DOI: 10.1016/j.caeo.2023.100142. URL: <https://www.sciencedirect.com/science/article/pii/S2666557323000204> (visited on 09/07/2025).
- Shannon, C E (1948). 'A Mathematical Theory of Communication'. In: *The Bell System Technical Journal* 27, pp. 379–423, 623–656.
- Shen, Shuanghong et al. (2024). 'A Survey of Knowledge Tracing: Models, Variants, and Applications'. In: *IEEE Transactions on Learning Technologies* 17, pp. 1858–1879. ISSN: 1939-1382, 2372-0050. DOI: 10.1109/TLT.2024.3383325. arXiv: 2105.15106[cs]. URL: <http://arxiv.org/abs/2105.15106> (visited on 08/08/2025).
- Srce, Sveučilišni računski centar- (2025). 'How (not) to design a L1 test vocabulary based on a lexical decision task: evidence from Croatian'. In: URL: <https://www.croris.hr/crosbi/publikacija/prilog-skup/885833> (visited on 16/06/2025).
- Varela, Fancisco J., Evan Thompson and Eleanor Rosch (1991). *The Embodied Mind*. MIT Press. Cambridge, Massachusetts. 308 pp.
- Wang, Zexuan, Ismaël Rafai and Marc Willinger (2023). 'Does age affect the relation between risk and time preferences? Evidence from a representative sample'. In: *Southern Economic Journal* 90.2, p. 341. DOI: 10.1002/soej.12662. URL: <https://hal.inrae.fr/hal-04217414> (visited on 06/09/2025).

Wechsler, David (2025). *Wechsler Adult Intelligence Scale—Third Edition*. URL: <https://psycnet.apa.org/record/9999-49755-000?doi=1> (visited on 14/01/2025).

William J., Clancey (1983). 'GUIDON'. In: *Journal of Computer-Based Instruction* 10.1, pp. 6–13. DOI: 10.5555/28590. URL: <https://billclancey.name/GUIDON-Clancey-CBI1982.pdf>.