

THE WELSH LANGUAGE IN THE DIGITAL AGE
Y GYMRAEG YN YR OES DDIGIDOL

Jeremy Evas



White Paper Series

Cyfres Papurau Gwyn

THE WELSH LANGUAGE IN THE DIGITAL AGE
Y GYMRAEG YN YR OES DDIGIDOL

Jeremy Evas

Ysgol y Gymraeg, Prifysgol Caerdydd
School of Welsh, Cardiff University

Georg Rehm, Hans Uszkoreit
(Golygyddion, editors)

Editors

Georg Rehm
DFKI
Alt-Moabit 91c
Berlin 10559
Germany
e-mail: georg.rehm@dfki.de

Hans Uszkoreit
DFKI
Alt-Moabit 91c
Berlin 10559
Germany
e-mail: hans.uszkoreit@dfki.de

ISSN 2194-1416 ISSN 2194-1424 (electronic)
ISBN 978-3-642-45371-7 ISBN 978-3-642-45372-4 (eBook)
DOI 10.1007/978-3-642-45372-4
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014931205

© Springer-Verlag Berlin Heidelberg 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)



RHAGARWEINIAD

Mae'r papur gwyn hwn yn rhan o gyfres sy'n hyrwyddo gwybodaeth am dechnoleg iaith a'i photensial. Mae'n targedu addysgwyr, newyddiadurwyr, llunwyr polisi, gwleidyddion, cynrychiolwyr a chymunedau ieithyddol ac eraill. Mae faint o dechnoleg iaith sydd ar gael yn Ewrop a'r defnydd ohoni yn amrywio rhwng ieithoedd. O ganlyniad, mae'r camau gweithredu y mae eu hangen i gefnogi ymchwil bellach ac i ddatblygu technolegau iaith hefyd yn wahanol ar gyfer pob iaith. Mae'r camau gweithredu angenrheidiol yn dibynnu ar sawl ffactor, megis cymhlethdod yr iaith a nifer ei siaradwyr. Yn y gyfres hon o bapurau gwyn, mae META-NET, rhwydwaith o ragoriaeth a ariennir gan y Comisiwn Ewropeaidd, wedi cynnal dadansoddiad o'r adnoddau a thechnolegau ieithyddol cyfredol sydd ar gael. Canolbwyntiodd y dadansoddiad ar 23 iaith swyddogol Ewrop yn ogystal ag ieithoedd cenedlaethol a rhanbarthol pwysig eraill Ewrop. Mae canlyniadau'r dadansoddiad hwn yn awgrymu bod llawer o fylchau ymchwil sylweddol ar gyfer pob iaith. Bydd dadansoddiad mwy manwl-arbenigol ac asesiad o'r sefyllfa bresennol yn helpu i fanteisio i'r eithaf ar ymchwil ac i leihau risgiau. Fis Tachwedd 2011, roedd META-NET yn cynnwys 54 o ganolfannau ymchwil o 33 o wledydd Ewropeaidd sydd yn gweithio gyda rhanddeiliaid o'r economi (cwmnïau meddalwedd, darparwyr a defnyddwyr technoleg) asiantaethau llywodraethol, sefydliadau ymchwil, sefydliadau anllywodraethol, cymunedau ieithyddol a Phrifysgolion Ewropeaidd. Ar y cyd a'r cymunedau hyn, mae META-NET wrthi'n creu gweledigaeth gyffredin ar gyfer technoleg ac agenda ymchwil strategol ar gyfer Ewrop amlieithog yn 2020

PREFACE

This white paper is part of a series that promotes knowledge about language technology and its potential. It addresses journalists, politicians, language communities, educators and others. The availability and use of language technology in Europe varies between languages. Consequently, the actions that are required to further support research and development of language technologies also differ. The required actions depend on many factors, such as the complexity of a given language and the size of its community.

META-NET, a Network of Excellence funded by the European Commission, has conducted an analysis of current language resources and technologies in this white paper series. The analysis focuses on the 23 official European languages as well as other important national and regional languages in Europe. The results of this analysis suggest that there are tremendous deficits in technology support and significant research gaps for each language. The given detailed expert analysis and assessment of the current situation will help maximise the impact of future research.

As of November 2011, META-NET consists of 54 research centres in 33 European countries. META-NET is working with stakeholders from economy (software companies, technology providers and users), government agencies, research organisations, non-governmental organisations, language communities and European universities. Together with these communities, META-NET is creating a common technology vision and strategic research agenda for multilingual Europe 2020.

Mae awdur y ddogfen hon yn ddiolchgar i awduron y papur gwyn ar yr Almaeneg [1] am ganiatâd i aildefnyddio peth deunydd annibynnol-ar-iaith o'u dogfen ac i Laura Amey, John Judge, Jonathan Morris, Michael Speed, a Colin Williams am eu cymorth gyda'r papur hwn.

Ariannwyd datblygu'r papur gwyn hwn gan Raglen Seithfed Fframwaith a Rhaglen Cefnogi Polisi TG y Comisiwn Ewropeaidd o dan y contractau T4ME (Cytundeb Grant 249119), CESAR (Cytundeb Grant 271022), META-NET4U (Cytundeb Grant 270893) a META-NORD (Cytundeb Grant 270899).

The author of this document is grateful to the authors of the white paper on German for permission to re-use selected language-independent materials from their document [1], and to Laura Amey, John Judge, Jonathan Morris, Michael Speed, and Colin Williams for their help with this paper.

The development of this white paper has been funded by the Seventh Framework Programme and the ICT Policy Support Programme of the European Commission under the contracts T4ME (Grant Agreement 249119), CESAR (Grant Agreement 271022), METANET4U (Grant Agreement 270893) and META-NORD (Grant Agreement 270899).



TABL CYNNWYS CONTENTS

Y Gymraeg yn yr Oes Ddigidol

| | | |
|----------|----------------------------------------------------------------------------|-----------|
| 1 | Crynodeb Gweithredol | 1 |
| 2 | Ieithoedd mewn perygl: her i Dechnoleg Iaith | 3 |
| 2.1 | Mae ffiniau ieithyddol yn llesteirio Cymdeithas Gwybodaeth Ewrop | 4 |
| 2.2 | Ein Hieithoedd mewn Perygl | 5 |
| 2.3 | Mae Technoleg Iaith yn dechnoleg allweddol ar gyfer galluogi | 5 |
| 2.4 | Cyfleoedd ar gyfer Technoleg Iaith | 6 |
| 2.5 | Caffael iaith mewn Bodau Dynol a Pheiriannau | 7 |
| 3 | Y Gymraeg yn y Gymdeithas Gwybodaeth Ewropeaidd | 9 |
| 3.1 | Ffeithiau cyffredinol | 9 |
| 3.2 | Nodweddion y Gymraeg | 10 |
| 3.3 | Datblygiadau diweddar | 13 |
| 3.4 | Hyrwyddo a Rheoleiddio Iaith | 14 |
| 3.5 | Hyfforddiant iaith a Thechnoleg ym myd Addysg | 16 |
| 3.6 | Agweddau Rhyngwladol | 16 |
| 3.7 | Y Gymraeg ar y Rhyngrwyd | 17 |
| 4 | Cymorth Technoleg iaith ar gyfer y Gymraeg | 18 |
| 4.1 | Pensaernïaeth Rhaglenni | 18 |
| 4.2 | Meysydd Rhaglen Craidd | 19 |
| 4.3 | Corpora | 28 |
| 4.4 | Argaeledd Offer ac Adnoddau | 29 |
| 4.5 | Cymhariaeth Drawsieithyddol | 29 |
| 4.6 | Casgliadau | 30 |
| 5 | Ynglŷn â META-NET | 34 |

THE WELSH LANGUAGE IN THE DIGITAL AGE

| | | |
|----------|-----------------------------------------------------------------------|-----------|
| 1 | Executive Summary | 35 |
| 2 | Languages at risk: a challenge for Language Technology | 37 |
| 2.1 | Language Borders hold back the European Information Society | 38 |
| 2.2 | Our Languages at Risk | 38 |
| 2.3 | Language Technology is a key enabling technology | 39 |
| 2.4 | Opportunities for language technology | 40 |
| 2.5 | Language Acquisition in Humans and Machines | 41 |
| 3 | The Welsh Language in the European Information Society | 43 |
| 3.1 | General Facts | 43 |
| 3.2 | Particularities of the Welsh Language | 44 |
| 3.3 | Recent developments | 46 |
| 3.4 | Language Promotion and Regulation | 48 |
| 3.5 | Language and Technology in Education | 50 |
| 3.6 | International Aspects | 50 |
| 3.7 | Welsh on the Internet | 50 |
| 4 | Language Technology Support for Welsh | 52 |
| 4.1 | Application Architectures | 52 |
| 4.2 | Core Application Areas | 53 |
| 4.3 | Translation Workflow and Content Management Systems | 59 |
| 4.4 | Availability of Tools and Resources for Welsh | 62 |
| 4.5 | Cross-language Comparison | 62 |
| 4.6 | Conclusions | 64 |
| 5 | About META-NET | 67 |
| A | Cyfeiriadau – References | 68 |
| B | Aelodau META-NET – META-NET Members | 71 |
| C | Cyfes Papurau Gwyn META-NET – The META-NET White Paper Series | 74 |

CRYNODEB GWEITHREDOL

Iaith yw'r prif ddull o gyfathrebu rhwng pobl. Mae'n ein galluogi i fynegi syniadau a theimladau, yn ein helpu i ddysgu ac addysgu, mae'n hanfodol ar gyfer byw. Dyma'r prif ddull o drosglwyddo diwylliant, ac mae'n symbol o'n hunaniaeth. O ystyried y globaleiddio sydd ohoni, mae gennym lawer o ddulliau o gyfathrebu'n hawdd gyda phobl o bob cwr o'r byd. Er enghraifft, mae'r technolegau gwybodaeth a chyfathrebu newydd wedi galluogi datblygu rhwydweithiau cymdeithasol sydd wedi annoga gwella rhyngweithio rhwng pobl o bron pob gwlad a diwylliant. Hefyd, yn ystod y blynyddoedd diwethaf, yr ydym wedi gweld symudiadau mawr o bobl tramor rhwng ein gwledydd, h.y. twristiaeth neu fewnfudo—sy'n creu'r angen am gyfathrebu ymysg gwahanol ieithoedd. Yn aml bydd y broblem cyfathrebu traws-ieithog hon yn cael ei datrys drwy ddefnyddio lingua franca. Yng ngwledydd Ewrop ceir enghraifft eglur o amrywiaeth ieithyddol a diwyllianol er gwaethaf y ffaith i Ewrop ymffurfio fwyfwy yn endid gwleidyddol ac economaidd penodol yn ystod y 60 mlynedd diwethaf. O ganlyniad, mae'n anochel ein bod yn wynebu heriau ieithyddol mewn bywyd bob dydd yn ogystal ag ym meysydd busnes, gwleidyddiaeth a'r gwyddorau.

Mae'n anochel ein bod yn wynebu heriau ieithyddol mewn bywyd bob dydd yn ogystal ag ym meysydd busnes, gwleidyddiaeth a'r gwyddorau

Bydd sefydliadau'r Undeb Ewropeaidd yn gwario tua biliwn Ewro y flwyddyn ar gynnal eu polisi amlieithrwydd, h.y. cyfieithu testunau a chyfieithu ar y pryd.

Ar yr un pryd, mae Saesneg yn troi'n lingua franca wrth i'r sefydliadau Ewropeaidd gyfathrebu â'u dinasyddion. Yn y DU, fel enghraifft o hyn, mae gennym sefyllfa debyg. Gan fod llawer o wasanaethau cyhoeddus bellach yn cael eu darparu naill ai'n uniongyrchol neu'n anuniongyrchol drwy ddulliau technolegol, mae darparu a chofnodi dewis iaith, a'r dechnoleg iaith angenrheidiol i roi'r dewis hwn ar waith bellach yn fater o bwys am sawl rheswm. Mae'r cyfiawnhad mwyaf perthnasol yn ymwneud â hyrwyddo: dinasyddiaeth weithgar, tegwch o ran mynediad at wasanaethau meddygol, hygyrchedd ar gyfer y rhai, er enghraifft, sydd â nam ar eu golwg, a chynrychiolaeth ddemocrataidd ei hun.

Gall technoleg bontio rhwng gwahanol ieithoedd

O'i chyfuno â dyfeisiau deallus a rhaglenni, bydd technoleg iaith y dyfodol yn gallu helpu dinasyddion i siarad yn rhwydd â'i gilydd a gwneud busnes â'i gilydd, hyd yn oed os nad ydynt yn siarad iaith gyffredin. Mae hyn, yng nghyd-destun deddfwriaeth ieithyddol sydd newydd ei phasio sy'n effeithio ar y gwasanaethau cyhoeddus yng Nghymru, yn hollbwysig. Bydd atebion technoleg iaith yn y pen draw yn gwasanaethu fel pont unigryw rhwng gwahanol ieithoedd. Fodd bynnag, mae'r technolegau iaith ac offer prosesu lleferydd sydd ar gael ar y farchnad ar hyn o bryd (yn amrywio o systemau ateb cwestiwn i ryngwynebau iaith naturiol, gan gynnwys systemau cyfieithu ac offer crynhoi, ymhlith llawer o rai eraill), yn dal i fod heb gyrraedd y nod uchelgeisiol hwn. Mor gynnar â diwedd y 1970au, sylweddolodd yr Undeb Ewropeaidd berthnasedd arwyddocaol technoleg iaith fel grym

i yrru undod Ewropeaidd, a dechrau ariannu'r prosiectau ymchwil cyntaf o fewn y maes datblygol hwn. Ar yr un pryd, sefydlwyd prosiectau cenedlaethol ac awtonomig a oedd yn cynhyrchu canlyniadau gwerthfawr ond ni chafwyd gweithgarwch Ewropeaidd cydlynol. Mae prif dechnolegau iaith heddiw yn dibynnu ar ddulliau ystadegol nad ydynt yn fanwl gywir, nad ydynt yn gwneud defnydd o ddulliau ieithyddol dyfnach, rheolau a gwybodaeth. Er enghraifft, bydd brawddegau'n cael eu cyfeithu'n awtomatig gan gymharu brawddeg newydd yn erbyn miloedd o frawddegau a gyfieithwyd yn flaenrol gan fodau dynol. Mae ansawdd yr allbwn i raddau helaeth yn dibynnu ar faint ac ansawdd y corpws sampl sydd ar gael. Fodd bynnag, mae hyd yn oed y dull ystadegol lled-anghywir yn llawer mwy cynhyrchiol na llafur cyfieithydd unigol nad yw'n ymelwa ar dechnoleg o'r fath, neu'n ymelwa ar rannu amser-go-iawn â gwaith cyfieithwyr eraill drwy Gof Cyfieithu. Er y gall cyfieithu brawddegau syml yn awtomatig mewn ieithoedd a chanddynt symiau digonol o ddeunydd testun ddarparu canlyniadau defnyddiol, mae dulliau ystadegol bas yn rhwym o fethu yn achos ieithoedd a chanddynt gorff llawer llai o ddeunydd sampl, neu yn achos brawddegau a chanddynt strwythurau cymhleth. Dadansoddi priodweddau strwythurol dyfnach ieithoedd yw'r unig ffordd ymlaen os dymunir adeiladu rhaglenni sy'n perfformio'n dda ar draws ystod eang o ieithoedd. Felly, yr ateb i broblem cyfathrebu rhwng ieithoedd yw adeiladu technolegau galluogi allweddol. Er mwyn cyrraedd y nod hwn a chadw amrywiaeth ddiwylliannol ac ieithyddol Ewrop, yn gyntaf, mae angen dadansoddiad systematig o nodweddion ieithyddol holl ieithoedd Ewrop, a chyflwr presennol technoleg iaith i'w cefnogi. Dyma bwrpas y papur gwyn hwn sy'n ymwneud â'r Gymraeg.

Yr ateb i broblem cyfathrebu rhwng ieithoedd yw
adeiladu technolegau galluogi allweddol

Fel y mae'r gyfres hon o bapurau gwyn yn ei ddangos, mae gwahaniaeth dramatig rhwng aelod-wladwriaethau

Ewrop o ran aeddfedrwydd y gwaith ymchwil a pharodrwydd yr ymateb i adnabod a gweithredu atebion iaith. Un o gynigion a chasgliadau sy'n deillio o'r dystiolaeth o'r hyn sydd ar gael yw bod y Gymraeg yn un o ieithoedd yr UE sydd angen ymchwil bellach cyn bod technoleg iaith yn barod i'w defnyddio mewn sefyllfaoedd bob dydd yn eang, cyn bod yr iaith wedi'i normaleiddio ym myd technoleg a chyn bod y dechnoleg honno'n normaleiddio'r Gymraeg i'w llawn botensial. Tra bo technoleg iaith yn dechnoleg galluogi ac nid yn nod ynddi ei hun ar gyfer y person-yn-y-stryd, mae'n hanfodol bod ieithoedd llai fel y Gymraeg yn derbyn sylw dyledus neu bydd eu siaradwyr yn cael eu difreinio ymhellach.

Mae'n hanfodol bod ieithoedd lleiafrifol yn cael
sylw dyledus, neu bydd eu siaradwyr yn cael eu
difreinio ymhellach

Mewn ardal lle na fydd yw siaradwyr iaith yn dod i adnabod ei gilydd mewn modd ethnig neu fel arall, mae gan dechnoleg rôl fawr i'w chwarae o ran cofnodi dewis iaith dinasyddion. Byddai'r sefyllfa galluogi ddelfrydol hon yn gweld y dechnoleg yn galluogi asiantau'r wladwriaeth a sectorau eraill i gynnig gwasanaethau yn Gymraeg *mewn modd rhagweithiol*, gan y bydd dewis iaith dinasyddion eisoes yn hysbys iddynt. Er bod y sefyllfa ddeddfwriaethol yng Nghymru a ddisgrifir isod yn datblygu ei disgwrs ar gyfer y wladwriaeth i fod yn ddarparwr gwasanaethau o'r fath, bydd technoleg yn galluogi tegwch o ran darpariaeth iaith i bob dinesydd (e.e. cyfeirio galwadau ffôn Cymraeg yn awtomatig i staff sy'n siarad Cymraeg, cyfateb staff Cymraeg yn y gwasanaethau cymdeithasol/meddygol â defnyddwyr gwasanaeth sy'n siarad Cymraeg ac yn y blaen).

Bydd technoleg yn galluogi tegwch o ran
darpariaeth iaith i bob dinesydd
