

THE IRISH AN
LANGUAGE IN GHAEILGE
THE DIGITAL SA RÉ
AGE DHIGITEACH

John Judge
Ailbhe Ní Chasaide
Rose Ní Dhubhda
Kevin P. Scannell
Elaine Uí Dhonnchadha



White Paper Series

Sraith Páipéar Bán

THE IRISH LANGUAGE IN THE DIGITAL AGE

AN
GHAEILGE
SA RÉ
DHIGITEACH

John Judge CNGI

Ailbhe Ní Chasaide Coláiste na Tríonóide

Rose Ní Dhubhda OÉ Gallimh

Kevin P. Scannell Saint Louis University

Elaine Uí Dhonnchadha Coláiste na
Tríonóide

Georg Rehm, Hans Uszkoreit
(eagarthóirí, 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 978-3-642-30557-3 ISBN 978-3-642-30558-0 (eBook)
DOI 10.1007/978-3-642-30558-0
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012942723

© Springer-Verlag Berlin Heidelberg 2012

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)



RÉAMHRÁ

PREFACE

Tá an páipéar bán seo mar chuid de shraith a chuireann chun cinn faisnéis maidir le teicneolaíocht teanga agus a cumas. Díríonn sé ar oideachasóirí, iriseoirí, polaiticoirí, pobail teanga agus daoine eile.

Tá difríochtaí idir infhaighteacht agus úsáid teicneolaíocht teanga san Eoraip ó theanga go teanga. Dá bhrí sin, ní hionann na gníomhaíochtaí a theastaíonn chun tuilleadh tacaíochta a thabhairt do thaighde agus forbairt teicneolaíochtaí teanga i ngach teanga. Braith-eann na gníomhaíochtaí a theastaíonn ar go leor tosca, cosúil le castacht na teanga agus méid a pobail.

Tá anailís déanta ag META-NET, Gréasán Sármhaitheasa arna mhaoiniú ag Coimisiún na hEorpa, ar na hacmhainní agus teicneolaíochtaí teanga reatha (lch. 87). Dhírigh an anailís seo ar na 23 teanga oifigiúil Eorpach mar aon le teangacha náisiúnta agus réigiúnacha tábhachtacha eile san Eoraip. Tugann torthaí na hanailíse seo le fios go bhfuil bearnaí taighde suntasacha i ngach teanga. Cuideoidh anailís agus measúnú sain-eolach níos mionsonraithe ar an gcás reatha le tionchar taighde bhreise a uasmhéadú agus na rioscaí a íoslaghdú.

Tá 54 ionad taighde as 33 tír i META-NET atá ag oibriú le geallsealbhóirí ó ghnólachtaí tráchtála (lch. 83), gníomhaireachtaí rialtais, tionscal, eagraíochtaí taighde, cuideachtaí bogearraí, soláthraithe teicneolaíochta agus ollscoileanna san Eoraip. Le chéile, tá siad ag cruthú comhfhís teicneolaíochta agus ag forbairt clár taighde straitéiseach a léiríonn an chaoi gur féidir le feidhmchláir teicneolaíochta teanga dul i ngleic le bearna i dtaighde faoi 2020.

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 (p. 87). 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 (p. 83). 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.

Tá údair na cáipéise seo buíoch d'údair an Pháipéir Bháin ar an nGearmáinis as cead a thabhairt ábhair neamhspleách ó theanga óna gcáipéis a athúsáid [1].

Mhainigh an Seachtú Creachlár agus Clár Tacaíochta Beartais TFC de Choimisiún na hEorpa faoi na conarthaí T4ME (Comhaontú Deontais 249 119), CESAR (Comhaontú Deontais 271 022), METANET4U (Comhaontú Deontais 270 893) agus META-NORD (Comhaontú Deontais 270 899) forbairt an pháipéir bháin seo.

The authors of this document are grateful to the authors of the white paper on German for permission to re-use selected language-independent materials from their document [1].

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 249 119), CESAR (Grant Agreement 271 022), METANET4U (Grant Agreement 270 893) and META-NORD (Grant Agreement 270 899).



CLÁR ÁBHAR CONTENTS

An GHAEILGE SA RÉ DHIGITEACH

1	Achoimre Fheidhmeach	1
2	Riosca dár dTeangacha agus Dúshlán dár dTeicneolaíocht Teanga	4
2.1	Cuireann Teorainneacha Teanga Bac ar Shochaí Faisnéise na hEorpa	5
2.2	Ár dTeangacha i mBaol	5
2.3	Is Teicneolaíocht Chumasaithe atá i dTeicneolaíocht Teanga	6
2.4	Deiseanna do Theicneolaíocht Teanga	6
2.5	Dúshlán le Teicneolaíocht Teanga	7
2.6	Éadail Teanga i nDaoine agus Innil	8
3	An Ghaeilge i Sochaí Faisnéise na hEorpa	10
3.1	Firicí Ginearálta	10
3.2	Sonraíochtaí na Gaeilge	11
3.3	Forbairtí Úrnua	14
3.4	Saothrú Teanga in Éirinn	15
3.5	Teanga san Oideachas	15
3.6	Gnéithe Idirnáisiúnta	15
3.7	An Ghaeilge ar an Idirlíon	16
4	Tacaíocht Teicneolaíochta Teanga don Ghaeilge	17
4.1	Ailtireachtaí Feidhmchláir	17
4.2	Croírimsí Feidhme	19
4.3	Réimsí Eile Feidhme	27
4.4	Tionscal agus Cláir TT	28
4.5	Taighde agus Oideachas TT in Éirinn	29
4.6	Infhaighteacht Uirlisí agus Acmhainní	31
4.7	Comparáid trasteanga	31
4.8	Conclúidí	34
5	Eolas faoi META-NET	39

THE IRISH LANGUAGE IN THE DIGITAL AGE

1	Executive Summary	41
2	Languages at Risk: a Challenge for Language Technology	43
2.1	Language Borders Hold back the European Information Society	44
2.2	Our Languages at Risk	44
2.3	Language Technology is a Key Enabling Technology	45
2.4	Opportunities for Language Technology	45
2.5	Challenges Facing Language Technology	46
2.6	Language Acquisition in Humans and Machines	46
3	The Irish Language in the European Information Society	48
3.1	General Facts	48
3.2	Particularities of the Irish Language	49
3.3	Recent Developments	52
3.4	Language Cultivation in Ireland	53
3.5	Language in Education	53
3.6	International Aspects	54
3.7	Irish on the Internet	54
4	Language Technology Support for Irish	56
4.1	Application Architectures	56
4.2	Core Application Areas	58
4.3	Other Application Areas	65
4.4	Language Technology Industry and Programmes	67
4.5	Language Technology Research and Education in Ireland	68
4.6	Availability of Tools and Resources	69
4.7	Cross-language comparison	70
4.8	Conclusions	72
5	About META-NET	77
A	Tagairtí – References	79
B	Baill de META-NET – META-NET Members	83
C	An Sraith Páipéar Bán META-NET – The META-NET White Paper Series	87

ACHOIMRE FHEIDHMEACH

Le 60 bliain anuas, tá an Eoraip tagtha chun cinn amháil struchtúr polaitiúil agus eacnamaíoch, ach tá éagsúlacht mhór fós inti maidir le cultúr agus teangacha. Dá bhrí sin, ón bPortaingéilis go Polainnis agus ó Iodáilis go hÍoslainnis, bíonn baic theanga i gceist le cumarsáid laethúil idir saoránaigh na hEorpa mar aon le cumarsáid sa saol gnó agus polaitiúil. Caitheann institiúidí an AE timpeall billiún euro sa bhliain ar a mbeartas ilteangachais a chothabháil, i. e. aistriú téacsanna agus teangaireacht a dhéanamh ar chumarsáid labhartha. An gá gur ualach chomh mór a bheadh anseo? Tig le teicneolaíocht teanga (TT) nua-aimseartha agus taighde theangeolaíoch cabhrú leis na teorainneacha teangeolaíoch a shárú. Nuair a nasctar an teicneolaíocht teanga le gairis agus feidhmchláir chliste, beidh sí in ann amach anseo cuidiú le muintir na hEorpa labhairt go héasca lena chéile agus gnó a dhéanamh lena chéile, fiú mura labhraíonn siad an teanga chéanna.

Cruthaíonn teicneolaíocht teanga droichid do thodhchaí na hEorpa.

Baineann geilleagar na hÉireann an-bhuntaíste as an margadh Eorpach aonair:

In 2010, b'ionann trádáil san AE agus 57.9% d'easpórtálacha na hÉireann agus b'ionann trádáil le tíortha Eorpacha eile agus 4.9% [2]. Ach tig le baic theanga stop a chur le gnó, go háirithe le SMEanna nach bhfuil na hacmhainní airgeadais acu an cás a chur ina cheart. An t-aon rogha (tubaisteach) a bheadh ann ar an Eoraip ilteangach seo go mbeadh teanga amháin

ceannasach agus go dtiocfadh sí in áit na dteangacha eile.

Bealach amháin chun bac teanga a shárú is ea teangacha iasachta a fhoghlaim. Ach gan tacaíocht theicneolaíoch, ní fhéadfadh saoránaigh na hEorpa agus a geilleagar, díospóireacht pholaitiúil agus dul chun cinn eolaíochta máistreacht a fháil ar na 23 teanga oifigiúil agus na 60 eile teanga Eorpach.

Is é réiteach na faidhbe príomhtheicneolaíochtaí cumasaithe a thógáil. Gheobhaidh rannpháirtithe Eorpacha buntáistí ollmhóra astu seo, ní hamháin sa chómhargadh Eorpach agus i gcaidreamh trádála le geilleagar forbarthacha, go háirithe le geilleagar éirith-eacha. Chun an sprioc seo a bhaint amach agus éagsúlacht chultúir agus theangeolaíoch na hEorpa a chaomhnú, is gá ar dtús anailís chórasach a dhéanamh ar mhionsonraíochtaí teanga na dteangacha Eorpacha ar fad, agus an bhail reatha atá ar thacaíocht teicneolaíochta dóibh. Beidh réitigh teicneolaíochta teanga ina ndroichid uathúla idir teangacha Eorpacha faoi dheireadh.

Teicneolaíocht teanga ina heochair don todhchaí.

Ní shásaíonn na huirlisí aistriúcháin uathoibríochas agus próiseála urlabhra atá ar an margadh faoi láthair an sprioc uailmhianach seo. Tá na rannpháirtithe ceannasacha sa réimse seo faoi úinéireacht phríobháideach go príomha chun brabach a dhéanamh agus bunaithe i Meiriceá Thuaidh. Bhí a fhios ag an AE faoi dheireadh na 1970idí go raibh teicneolaíocht teanga anábhartha

mar thiománaí d'aontacht Eorpach, agus chuir siad tús leis na chéad tionscadail taighde, amhail EUROTRA. Ag an am céanna, bunaíodh go leor tionscadail náisiúnta a chothaigh torthaí luachmhara, ach nár eascair gníomhaíocht Eorpach astu. I gcodarsnacht leis an iarracht mhaoinithe ardroghnach seo, bhunaigh sochaithe ilteangacha eile cosúil leis an India (22 teanga oifigiúil) agus an Afraic Theas (11 theanga oifigiúla) cláir náisiúnta fhadtéarmacha do thaighde agus forbairt teicneolaíochta teanga.

Tá na gníomhairí is mó i dteicneolaíocht teanga inniu ag brath ar chur chuige staitistiúil neamhbheacht nach n-úsáideann modhanna agus faisnéis theangeolaíoch níos doimhne. Mar shampla, aistrítear abairtí go huathoibríoch trí abairt nua a chur i gcomparáid leis na mílte abairt a d'aistrigh daoine cheana. Braitheann caighdeán an aschuir go mór ar mhéid agus caighdeán an chorpais shamplaigh atá ar fáil. Cé gur féidir torthaí sásúla a fháil as aistriú uathoibríoch abairtí simplí i dteangacha ina bhfuil dóthain ábhair théacs ar fáil, ní éireoidh le modhanna staitistiúla chomh héadomhain sin i gcás teangacha ina bhfuil corpas ábhar téacs i bhfad níos lú nó i gcás abairtí le struchtúir chasta.

Cuidíonn Teicneolaíocht Teanga chun an Eoraip a chomhaontú.

Dá bhrí sin, tá cinneadh déanta ag an Aontas Eorpach maoniú a thabhairt do thionscadail cosúil le EuroMatrix agus EuroMatrixPlus (ó 2006) agus iTranslate4 (ó 2010) a dhéanann taighde bunúsach agus feidhmeach agus a chothaíonn acmhainní chun réitigh theicneolaíochta teanga ardchaighdeán a bhunú do gach teanga Eorpach. Is é an t-aon bhealach chun dul chun cinn a dhéanamh anailís a dhéanamh ar airí struchtúrtha níos doimhne teangacha má theastaíonn uainn feidhmchláir a thógáil a fheidhmíonn go maith don réimse teangacha ar fad san Eoraip.

Tá rath bainte amach ag taighde Eorpach sa réimse seo cheana. Mar shampla, úsáidtear MOSES, bogearra aistriúcháin uathoibríoch foinse oscailte, a forbraíodh go príomha trí thionscadail taighde Eorpacha anois i seirbhísí aistriúcháin an Aontais Eorpaigh. Chuir tionscadal Verbmobil, arna mhaoiniú ag Aireacht Oideachais agus Taighde na Gearmáine (BMBF) idir 1993 agus 2000, an Ghearmáin chun tosaigh i réimse an taighde um aistriú urlabhra ar feadh scathaimh. Dúnadh nó bogadh chuig ionad eile go leor de na saotharlanna taighde agus forbartha a bhí sa Ghearmáin ag an am (e.g., IBM agus Philips). In ionad dul ar aghaidh ó thorthaí na dtionscadal taighde seo, dhírigh an Eoraip ar ghníomhaíochtaí taighde aonair nach raibh tionchar chomh forleatach acu ar an margadh. Is féidir luach geilleagrach na n-iarrachtaí is luaithe fiú a fheiceáil sa líon seachthionscadal. Díoladh an chuideachta Trados, a bunaíodh i 1984, le SDL atá bunaithe sa RA in 2005.

Ag tarraingt ar an léargas a fuarthas go dtí seo, is léir go mbeidh teicneolaíocht teanga 'hibride' an lae inniu, a mheascann próiseáil dhomhain le modhanna staitistiúla, in ann an bhearna idir gach teanga Eorpach agus teangacha eile nach iad a líonadh. Mar a léirítear sa tsraith páipéar bána seo, tá difríocht ollmhór sa staid ullmhachta maidir le réitigh theanga agus an staid taighde idir ballstáit na hEorpa. Tá an Ghacilge, mionteanga san AE, agus go deimhin in Éirinn, i mbaol bheith fágtha taobh thiar i ndáil le dul chun cinn mura ngníomhaítear i dtreo bunteicneolaíocht teanga comhpháirte a sholáthar le tacú leis an teanga. Ag an am céanna, tá Éire i staid láidir chun na teicneolaíochtaí seo a fhorbairt agus ról lárnach bheith aici go geilleagrach agus i dtaobh na heolaíochta i ndul chun cinn suntasach a dhéanamh sa réimse ina iomláine, de bharr ardchaighdeán in aistriúcháin agus logánú agus taighde teicneolaíocht teanga.

Is í sprioc fhadtéarmach META-NET teicneolaíocht teanga ardchaighdeán a chur ar fáil do gach teanga d'fhonn comhaontacht pholaitiúil agus geilleagrach