META-NET

White Paper Series

Serje ta' White Papers

THE MALTESE LANGUAGE IN THE DIGITAL AGE

IL-LINGWA MALTIJA FL-ERA DIĠITALI

Mike Rosner
Jan Joachimsen

Springer

White Paper Series

Serje ta' White Papers

THE MALTESE LANGUAGE IN THE DIGITAL AGE

IL-LINGWA MALTIJA FL-ERA DIĠITALI

Mike Rosner University of Malta
Jan Joachimsen University of Malta

Georg Rehm, Hans Uszkoreit
(edituri, 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 978-3-642-30680-8 ISBN 978-3-642-30681-5 (eBook)
DOI 10.1007/978-3-642-30681-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012945122

© 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)

DAĦLA

Din il-White Paper hija għall-edukaturi, ġurnalisti, politikanti, kommunitajiet ta' lingwi u oħrajn, li jixtiequ jistabbilixxu Ewropa tabilħaqq multilingwali.

Din hija parti minn serje White Papers li jippromwovu għarfien dwar it-teknoloġija lingwistika u l-potenzjal tagħha. Id-disponibiltà u l-użu tat-teknoloġija lingwistika fl-Ewropa tvarja bejn lingwa u oħra. Konsegwentement, l-azzjonijiet li huma meħtieġa sabiex jiġu appoġġjati r-riċerka u l-iżvilupp tat-teknoloġiji lingwistiċi ivarjaw ukoll f'kull lingwa. L-azzjonijiet meħtieġa jiddependu fuq bosta fatturi, bħal kumplessità ta' lingwa partikolari u d-daqs tal-komunità tagħha.

META-NET, Netwerk ta' Eċċellenza tal-Kummissjoni Ewropea, wettaq analiżi dwar ir-riżorsi u t-teknoloġiji lingwistiċi kurrenti f'din is-serje ta' *white papers* (p. 91). Din l-analiżi kienet ibbażata fuq 23 lingwa uffiċjali Ewropeja, kif ukoll lingwi reġjonali oħra importanti fl-Ewropa. Ir-riżultati ta' dan l-analiżi jissuġġerixxu li hemm bosta nuqqasijiet fir-riċerka għal kull lingwa. Analiżi aktar dettaljata u esperta u assessjar tas-sitwazzjoni kurrenti għandha tgħin sabiex timmassimizza l-impatt ta' riċerka addizzjonali.

Minn Novembru 2011 META-NET tikkonsisti f'54 ċentru ta' riċerka minn 33 pajjiż [1] (p. 87) li qed jaħdmu ma' partijiet interessati minn negozji kummerċjali, aġenziji governattivi, industriji, organizzazzjonijiet ta' riċerka, kumpaniji ta' software, fornituri ta' teknoloġija u universitajiet Ewropej. Flimkien, dawn qed joħolqu viżjoni teknoloġika komuni filwaqt li jiżviluppaw aġenda ta' riċerka strateġika li turi kif applikazzjonijiet teknoloġiċi lingwistiċi jistgħu jindirizzaw xi nuqqasijiet ta' riċerka sal-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 differs. 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. 91). The analysis focused 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 additional research.

META-NET consists of 54 research centres from 33 European countries [1] (p. 87). META-NET is working with stakeholders from economy (software companies, technology providers, 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.

META-NET – office@meta-net.eu – http://www.meta-net.eu

L-awturi ta' dan id-dokument huma grati lejn l-awturi tal-*White Paper* Ġermaniż għall-permess biex jużaw mill-ġdid materjali magħżulin mid-dokument tagħhom [2] li huma in-dipendenti mill-lingwa.

Grazzi lil Ritianne Stanyer u Roberta Abela għat-translazzjoni ta' dan id-dokument. L-awturi huma obbligati lejn il-Professur Ray Fabri, li l-artiklu [3] tiegħu kien l-ispirazzjoni għal ħafna mill-kontenut u ħafna mill-eżempji f'din id-dokument.

L-iżvilupp ta' din il-white paper ġie ffinanzjat mis-Seba' Programm ta' Qafas u l-Programm ta' Appoġġ għall-Politika dwar l-ICT tal-Kummissjoni Ewropea taħt il-kuntratti T4ME (Grant Agreement 249 119), CESAR (Grant Agreement 271 022), METANET4U (Grant Agreement 270 893) u META-NORD (Grant Agreement 270 899).

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 [2].

Thanks to Ritianne Stanyer and Roberta Abela for the translation of this document. The authors are indebted to Prof. Ray Fabri, whose article [3] was the inspiration for much of the content and many of the examples in the document.

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).

WERREJ CONTENTS

IL-LINGWA MALTIJA FL-ERA DIĠITALI

1 Sommarju Eżekuttiv 1

2 Riskju għal-Lingwi Tagħna u Sfida għat-Teknoloġija Lingwistika 4
2.1 Il-Konfini Lingwistiċi jfixklu s-Soċjetà Ewropea tal-Informazzjoni 5
2.2 Il-Lingwi Tagħna f'Riskju 5
2.3 It-Teknoloġija Lingwistika hija Teknoloġija Katalizzanti Ewlenija 6
2.4 Opportunitajiet għat-Teknoloġija Lingwistika 6
2.5 Sfidi li t-Teknoloġija Lingwistika Taffaċċja 7
2.6 Il-Ksib tal-Lingwi tal-bnedmin u tal-magni 8

3 Il-Malti fis-Soċjetà tal-Informazzjoni Ewropea 10
3.1 Fatti Ġenerali 10
3.2 Partikolaritajiet tal-Lingwa Maltija 11
3.3 Żviluppi riċenti 12
3.4 Il-Kultivazzjoni tal-Lingwa f'Malta 13
3.5 Il-Lingwi fl-Edukazzjoni 14
3.6 Aspetti internazzjonali 16
3.7 Il-Malti fuq l-Internet 17

4 Appoġġ ta' Teknoloġija Lingwistika għall-Malti 22
4.1 Arkitetturi ta' Applikazzjonijiet 22
4.2 Oqsma ewlenin ta' applikazzjoni 24
4.3 Oqsma oħra tal-applikazzjoni 32
4.4 Programmi tal-Edukazzjoni 33
4.5 Programmi u Sforzi Nazzjonali 35
4.6 Disponibbiltà ta' Għodod u Riżorsi 36
4.7 Tqabbil ta' Trans-Lingwi 38
4.8 Konklużjonijiet 39

5 Dwar META-NET 42

THE MALTESE LANGUAGE IN THE DIGITAL AGE

1 Executive Summary 43

2 Languages at Risk: a Challenge for Language Technology 46
2.1 Language Borders Hold back the European Information Society 47
2.2 Our Languages at Risk 47
2.3 Language Technology is a Key Enabling Technology 48
2.4 Opportunities for Language Technology 48
2.5 Challenges Facing Language Technology 49
2.6 Language Acquisition in Humans and Machines 49

3 The Maltese Language in the European Information Society 51
3.1 General Facts 51
3.2 Particularities of the Maltese Language 52
3.3 Recent Developments 54
3.4 Language Cultivation in Malta 55
3.5 Language in Education 56
3.6 International Aspects 57
3.7 Maltese on the Internet 58

4 Language Technology Support for Maltese 62
4.1 Application Architectures 62
4.2 Core Application Areas 63
4.3 Other Application Areas 71
4.4 Educational Programmes 72
4.5 National Projects and Initiatives 73
4.6 Availability of Tools and Resources 75
4.7 Cross-language comparison 76
4.8 Conclusions 77

5 About META-NET 80

A Referenzi – References 81

B Membri ta' META-NET – META-NET Members 87

C Is-Serje ta' White Papers ta' META-NET – The META-NET White Paper Series 91

SOMMARJU EŻEKUTTIV

Matul dawn l-aħħar sittin sena, l-Ewropa saret struttura politika u ekonomika distinta, iżda kulturalment u lingwistikament għadha diversa ħafna. Dan ifisser li mill-Portugiż għall-Pollakk u t-Taljan għall-Islandiż, ta' kuljum il-komunikazzjoni bejn iċ-ċittadini tal-Ewropa kif ukoll il-komunikazzjoni fl-oqsma tan-negozju u l-politika hija inevitabbilment ikkonfrontata minn ostakli lingwistiċi. L-istituzzjonijiet tal-UE jonfqu madwar biljun euro fis-sena fuq iż-żamma tal-politika tagħhom tal-multilingwiżmu, jiġifieri, it-traduzzjoni ta' testi u l-interpretar tal-komunikazzjoni mitkellma. Madankollu, għandu dan ikun ta' daqsekk piż? It-teknoloġija moderna tal-lingwi u r-riċerka lingwistika jistgħu jkunu ta' kontribuzzjoni sinifikanti biex jitwaqqgħu dawn l-ostakli lingwistiċi. Meta kkombinati mal-apparati u l-applikazzjonijiet intelliġenti, it-teknoloġija lingwistika tkun fil-futur tista' tgħin lill-Ewropej jitkellmu ma' xulxin b'mod faċli u jagħmlu n-negozju ma' xulxin anke jekk ma jkunux jitkellmu lingwa komuni.

It-teknoloġija lingwistika tibni rabtiet.

L-ostakli tal-lingwa jistgħu jwaqqfu n-negozju, b'mod speċjali għall-SMEs li m'għandhomx il-mezzi finanzjarji biex ireġġgħu lura s-sitwazzjoni. L-unika alternattiva (inkonċepibbli) għal din it-tip ta' Ewropa multilingwali tkun li tippermetti lingwa waħda biex tieħu pożizzjoni dominanti u tispiċċa tissostitwixxi kull lingwa oħra. Mod klassiku ta' kif tista' tegħleb l-ostaklu tal-lingwa huwa li titgħallem il-lingwi barranin. Iżda mingħajr appoġġ teknoloġiku, li tkun taf it-23 lingwa uffiċjali tal-Istati Membri tal-Unjoni Ewropea u xi sittin lingwa Ewropea oħra huwa ostaklu insormontabbli għaċ-ċittadini tal-Ewropa u l-ekonomija, id-dibattitu politiku, u l-progress xjentifiku tagħha.

Is-soluzzjoni hija li jinbnew teknoloġiji ewlenin li jippermettu dan. Dawn ikunu joffru lill-atturi Ewropej vantaġġi tremendi, mhux biss fis-suq komuni Ewropew iżda wkoll f'relazzjonijiet ta' kummerċ ma' pajjiżi terzi, b'mod speċjali l-ekonomiji emerġenti. Biex jintlaħaq dan il-għan u tiġi preservata d-diversità kulturali u lingwistika tal-Ewropa, huwa neċessarju li l-ewwel issir analiżi sistematika tal-partikolaritajiet tal-lingwi Ewropej kollha u l-istat attwali tal-appoġġ tat-teknoloġija lingwistika li hawn għalihom. Soluzzjonijiet tat-teknoloġiji lingwistiċi se jservu eventwalment bħala rabta unika bejn il-lingwi tal-Ewropa.

L-għodda ta' traduzzjoni awtomatizzata u tal-ipproċessar tad-diskors li huma attwalment disponibbli fis-suq għadhom pjuttost lura minn dan il-għan ambizzjuż. L-atturi dominanti f'dan il-qasam huma primarjament l-intrapriżi b'sidien privati għall-profitt ibbażati fl-Amerka ta' Fuq. Diġà fl-aħħar tal-1970, l-UE rrealizzat ir-relevanza profonda tat-teknoloġija lingwistika bħala sewwieq lejn l-unità Ewropea , u bdiet tiffinanzja l-ewwel proġetti ta' riċerka tagħha, bħall-EUROTRA. Fl-istess żmien, proġetti nazzjonali ġew imwaqqfa u ġġeneraw riżultati ta' valur iżda qatt ma wasslu għal azzjoni Ewropea kkonċertata. F'kuntrast ma' dan l-isforz ta' finanzjament ferm selettiv, soċjetajiet multilingwali oħra bħall-Indja (22 lingwa uffiċjali) u l-Afrika t'Isfel

(11-il lingwa uffiċjali) waqqfu reċentement programmi nazzjonali fit-tul għar-riċerka tal-lingwi u l-iżvilupp tat-teknoloġija.

L-atturi predominanti fit-teknoloġija lingwistika llum jiddependu fuq approċċi ta' statistika mhux preċiżi li ma jagħmlux użu minn metodi lingwistiċi u għarfien aktar fondi. Per eżempju, is-sentenzi jiġu awtomatikament tradotti billi tiġi kkumparata sentenza ġdida ma' eluf ta' sentenzi oħrajn li jkunu ġew tradotti qabel minn umani. Il-kwalità tar-riżultat il-biċċa l-kbira tiddependi fuq l-ammont u l-kwalità tal-kampjun tal-korp disponibbli. Filwaqt li t-traduzzjoni awtomatika ta' sentenzi sempliċi fil-lingwi b'ammonti suffiċjenti ta' materjal ta' testi disponibbli tista' tikseb riżultati utli, metodi ta' statistika vojta bħal dawn huma destinati li jfallu fil-każ ta' lingwi b'korp ta' kampjuni tal-materjal ferm iżgħar jew fil-każ ta' sentenzi bi strutturi kumplessi.

It-teknoloġija lingwistika bħala ċavetta għall-futur.

Għalhekk l-Unjoni Ewropea ddeċidiet li tiffinanzja proġetti bħall-EuroMatrix u l-EuroMatrixPlus (mill-2006) u iTranslate4 (mill-2010) li jwettqu riċerka bażika u applikata u jiġġeneraw riżorsi biex jiġu stabbiliti soluzzjonijiet għat-teknoloġija lingwistika ta' kwalità għolja għall-lingwi Ewropej kollha. Li tanalizza l-proprjetajiet strutturali aktar fondi tal-lingwi huwa l-uniku pass 'il quddiem jekk irridu nibnu applikazzjonijiet li jaħdmu tajjeb tul il-firxa kollha tal-lingwi tal-Ewropa.

Ir-riċerka Ewropea f'dan il-qasam diġà kisbet numru ta' suċċessi. Per eżempju, is-servizzi ta' traduzzjoni tal-Unjoni Ewropea issa jużaw is-softwer ta' traduzzjoni bil-magni b'sors miftuħ MOSES li ġie prinċipalment żviluppat permezz ta' proġetti ta' riċerka Ewropej. F'Malta, l-oqsma tat-teknoloġija lingwistika l-aktar avvanzati bħalissa huma dawk tas-sinteżi tat-taħdit u l-korpora tat-testi: fil-qasam tas-sinteżi tat-taħdit bil-Malti, proġett appoġġjat mill-Gvern parzjalment iffinanzjat mill-fond ta' żvilupp reġjonali tal-UE qiegħed fil-proċess li jwassal it-teknoloġija tat-taħdit lill-persuni b'diżabiltà. Il-konsorzju, li jikkonsisti f'SME (Crimson Wing Ltd), fondazzjoni (FITA, Fundazzjoni għall-Aċċess tat-TI), u l-Università, wiegħed li dawn ir-riżorsi se jkunu disponibbli għal skopijiet ta' riċerka. Fil-qasam tal-korpora tat-testi, is-Server għar-Riżorsi Lingwistiċi bil-Malti (MLRS) qed jagħti frott u sforzi sinifikanti li għadhom għaddejjin fl-Università, permezz tal-Istitut tal-Lingwistika (A. Gatt, C. Borg, R. Fabri) u d-Dipartiment ta' Sistemi Intelliġenti tal-Kompjuter (M. Rosner), li jsostnu u jiżviluppaw dan. Bħalissa il-korpus jinkludi madwar 100M kelma, u hemm aktar għodod ippjanati inkluż tagger għall-kategoriji tal-kliem u ċekkjatur ortografiku.

It-teknoloġija lingwistika tgħin biex tgħaqqad lill-Ewropa.

Jekk inħarsu lejn l-għarfien miksub s'issa, jidher li t-teknoloġija 'ibrida' tal-lingwi tal-lum li tħallat metodi ta' pproċessar fondi ma' dawk ta' statistika se tkun tista' timla' l-vojt ta' bejn il-lingwi Ewropej kollha u aktar. Bħalma din is-serje ta' *white papers* turi, hemm differenza drammatika fl-istat ta' prontezza fir-rigward ta' soluzzjonijiet ta' lingwi u l-istat tar-riċerka bejn l-Istati Membri tal-Ewropa. Din il-*white* paper għall-lingwa Maltija turi li hemm il-potenzjal għal industrija tat-teknoloġija lingwistika u ambjent tar-riċerka f'Malta. Iżda għalkemm numru ta' teknoloġiji u riżorsi jeżisti, hemm ħafna anqas minn lingwi Ewropej li huma "akbar" u ċertament mhux biżżejjed biex tiġi appoġġjata l-firxa kompleta ta' applikazzjonijiet sensittivi għall-lingwi li huma disponibbli għal dawk il-lingwi l-oħra.

Skont il-valutazzjoni ddettaljata f'dan ir-rapport, il-kisba ta' suċċess fit-teknoloġija tal-lingwa Maltija tirrekjedi ċiklu sħiħ ta' bidliet li jkun jinvolvi fornituri ta' kontenut, żviluppaturi u utenti tat-teknoloġija lingwistika.

Xi bidliet fil-politika tal-lingwa nazzjonali jridu jiġu implimentati qabel ma xi suċċessi għall-Lingwa Maltija jkunu jistgħu jinkisbu.

L-għan fit-tul tal-META-NET huwa li tiġi introdotta teknoloġija lingwistika bi kwalità għolja għall-lingwi kollha sabiex tinkiseb l-unità politika u ekonomika permezz tad-diversità kulturali. It-teknoloġija tkun tgħin biex tkisser l-ostakli eżistenti u jinbnew ir-rabtiet bejn il-lingwi tal-Ewropa. Dan jirrikjedi l-partijiet interessati kollha – fil-politika, ir-riċerka, in-negozju, u s-soċjetà – biex jingħaqdu l-isforzi tagħhom fil-futur.

Din is-serje ta' *white papers* tikkumplimenta azzjonijiet strateġiċi oħra meħuda minn META-NET (ara l-appendiċi għal ħarsa ġenerali). Informazzjoni aġġornata bħall-verżjoni attwali tal-*vision paper* [4] tal-META-NET jew l-Aġenda għar-Riċerka Strateġika (SRA) jistgħu jinstabu fuq il-websajt tal-META-NET http://www.meta-net.eu.