

THE CATALAN LANGUAGE
IN THE DIGITAL AGE

LA LLENGUA
CATALANA
A L'ERA
DIGITAL

Asunción Moreno
Núria Bel
Eva Revilla
Emília Garcia
Sisco Vallverdú



White Paper Series

Sèrie de Llibres Blancs

THE CATALAN
LANGUAGE
IN THE
DIGITAL AGE

LA LLENGUA
CATALANA
A L'ERA
DIGITAL

Asunción Moreno UPC

Núria Bel UPF

Eva Revilla UPF

Emília Garcia UPC

Sisco Vallverdú UPC

UPC Universitat Politècnica de Catalunya

UPF Universitat Pompeu Fabra

Georg Rehm, Hans Uszkoreit
(editors 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 978-3-642-30677-8 ISBN 978-3-642-30678-5 (eBook)
DOI 10.1007/978-3-642-30678-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012942724

© 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)



PRÒLEG

PREFACE

Aquest llibre blanc forma part d'una sèrie que vol donar a conèixer les tecnologies del llenguatge i el seu potencial. S'adreça a periodistes, polítics, comunitats lingüístiques i educadors entre d'altres. A Europa, la disponibilitat de tecnologies del llenguatge i el seu ús varia segons l'idioma. En conseqüència, les accions que es requereixen per reforçar la investigació i el desenvolupament d'aquestes tecnologies també són diferents per a cada idioma. Les accions necessàries depenen de molts factors, com ara la complexitat d'una llengua determinada o de la mida de la comunitat de parlants.

META-NET, una xarxa d'excel·lència finançada per la Comissió Europea, ha dut a terme una anàlisi de l'estat actual dels recursos i les tecnologies del llenguatge per a les 23 llengües oficials europees, així com per a d'altres idiomes importants a nivell nacional i regional a Europa. Els resultats d'aquesta anàlisi suggereixen que hi ha encara mancances significatives en la investigació necessària per a cada idioma. Una anàlisi més detallada i una avaluació experta de la situació actual han de contribuir a maximitzar l'impacte de noves investigacions. A novembre de 2011, META-NET es compon de 54 centres de recerca de 33 països (pàg. 71) que estan treballant amb tots els agents implicats: comercials, agències de govern, organitzacions de recerca, empreses de programari, proveïdors de tecnologia i universitats europees. Junts, estan treballant en una visió comuna de la tecnologia i desenvolupant una agenda estratègica de la recerca que mostra com les aplicacions de les tecnologies del llenguatge podran resoldre els problemes actuals abans de l'any 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. 75). 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.

As of November 2011, META-NET consists of 54 research centres from 33 European countries (p. 71). 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.

Els autors d'aquest document agraeixen als autors del Llibre Blanc en Alemany [1] el permís per reutilitzar fragments del seu document independents de la llengua.

El desenvolupament d'aquest document ha estat finançat pel Setè Programa Marc i el Programa de Suport a Polítiques de les TIC de la Comissió Europea amb els contractes T4ME (acord de subvenció 249 119), CESAR (acord de subvenció 271 022), METANET4U (acord de subvenció del 270 893) i META-NORD (acord de subvenció 270 899).

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



SUMARI CONTENTS

LA LENGUA CATALANA A L'ERA DIGITAL

1	Resum	1
2	Un risc per a les nostres llengües i un repte per a les tecnologies del llenguatge	3
2.1	Les fronteres lingüístiques dificulten la societat de la informació europea	4
2.2	Les nostres llengües en risc	4
2.3	Les tecnologies del llenguatge són una tecnologia clau	5
2.4	Oportunitats per a les tecnologies del llenguatge	6
2.5	Els reptes de les tecnologies del llenguatge	6
2.6	Adquisició de la llengua	7
3	El català a la societat de la informació europea	9
3.1	Aspectes generals	9
3.2	Particularitats del català	9
3.3	Desenvolupament recent	10
3.4	Promoció de la llengua catalana	11
3.5	La llengua a l'educació	11
3.6	Aspectes internacionals	12
3.7	El català a Internet	14
4	Suport de les tecnologies del llenguatge per al català	16
4.1	Arquitectures de les aplicacions de tecnologies del llenguatge	16
4.2	Àrees principals d'aplicació	17
4.3	Altres àrees d'aplicació	25
4.4	Les tecnologies del llenguatge a l'educació	26
4.5	Programes de suport per a les tecnologies del llenguatge	27
4.6	Eines i recursos disponibles	28
4.7	Comparació entre llengües	29
4.8	Conclusions	30
5	Quant a META-NET	34

THE CATALAN LANGUAGE IN THE DIGITAL AGE

1	Executive Summary	35
2	Risk for Our Languages and a Challenge for Language Technology	37
2.1	Language Borders Hinder the European Information Society	38
2.2	Our Languages at Risk	38
2.3	Language Technology is a Key Enabling Technology	38
2.4	Opportunities for Language Technology	39
2.5	Challenges Facing Language Technology	40
2.6	Language Acquisition in Humans and Machines	40
3	Catalan in the European Information Society	42
3.1	General Facts	42
3.2	Particularities of the Catalan Language	42
3.3	Recent Developments	43
3.4	Official Language Protection in Catalonia	44
3.5	Language in Education	44
3.6	International Aspects	45
3.7	Catalan on the Internet	47
4	Language Technology Support for Catalan	49
4.1	Application Architectures	49
4.2	Core Application Areas	50
4.3	Other Application Areas	57
4.4	Educational Programmes	58
4.5	National Projects and Efforts	59
4.6	Availability of Tools and Resources	60
4.7	Cross-language comparison	60
4.8	Conclusions	62
5	About META-NET	66
A	Referències – References	67
B	Membres de META-NET – META-NET Members	71
C	La sèrie de llibres blancs de META-NET – The META-NET White Paper Series	75

RESUM

El llenguatge és el mitjà principal de comunicació entre les persones. Ens permet expressar idees i sentiments, ens ajuda a aprendre i a ensenyar, és essencial per la vida, és el principal vehicle per transmetre la cultura i és un símbol de la nostra identitat cultural.

El llenguatge és el mitjà principal de comunicació entre les persones.

Atès el nivell de globalització de la nostra societat, han sorgit moltes formes de comunicar-nos fàcilment amb persones de tot el món. Per exemple, les noves tecnologies de la informació i la comunicació han permès el desenvolupament de xarxes socials que han fomentat i millorat la interacció entre persones de, virtualment, tots els països i cultures. També hem vist els últims anys grans moviments de persones que, per turisme o per migracions, viatgen a països estrangers, i pateixen dificultats de comunicació. La comunicació entre parlants de llengües diferents sovint es resol amb l'ús d'una llengua franca.

Els països d'Europa són un clar exemple de diversitat cultural i lingüística tot i que durant els últims seixanta anys, Europa ha anat convertint-se en una entitat política i econòmica ben definida. Això significa que, des del portuguès al polonès i des de l'italià fins al islandès, els ciutadans europeus es veuen inevitablement enfrontats amb desafiaments lingüístics també en la seva vida quotidiana, a més a més de en els ambients professionals, polítics i científics. Les institucions de la Unió Europea gasten al voltant de mil milions d'euros l'any per mantenir

les seves polítiques de multilingüisme, és a dir, per traduir textos i oferir traducció i interpretació simultània en situacions de comunicacions orals. Però, en paral·lel, l'anglès esdevé la llengua franca per a la comunicació entre els ciutadans europeus.

Un altre exemple d'aquest escenari multilingüe el veiem a Espanya, on hi ha una llengua oficial, l'espanyol, també anomenat castellà, i tres llengües co-oficials a diferents territoris: el català, el gallec i el basc. Preservar el multilingüisme a Espanya no ha estat una tasca fàcil, i ha estat el resultat de un complex procés de protecció de la identitat cultural dintre i entre les diferents regions i pobles d'Espanya. De manera semblant a l'ús de l'anglès al cas d'Europa, la comunicació entre els ciutadans de diferents àrees d'Espanya, tot sovint passa per l'ús del castellà com a llengua franca.

El multilingüisme representa una herència cultural que cal preservar.

En ambdós casos, l'europeu i l'espanyol, el multilingüisme és una herència cultural que ha de ser preservada. La globalització no hauria de ser un mecanisme que promogui l'abandonament de la nostra rica herència lingüística i cultural ni convidar-nos a abandonar l'ús de la nostra llengua materna a favor d'una altra. En un entorn de comunicació global, la societat ha de ser capaç de trobar maneres de comunicar-se amb la resta del món tot preservant la llengua dels ciutadans i, amb ella, la seva identitat cultural.