

THE ITALIAN LANGUAGE IN THE DIGITAL AGE
LA LINGUA ITALIANA NELL'ERA DIGITALE

Nicoletta Calzolari
Bernardo Magnini
Claudia Soria
Manuela Speranza



White Paper Series

Collana Libri Bianchi

THE ITALIAN
LANGUAGE IN
THE DIGITAL
AGE

LA LINGUA
ITALIANA
NELL'ERA
DIGITALE

Nicoletta Calzolari CNR-ILC

Bernardo Magnini FBK

Claudia Soria CNR-ILC

Manuela Speranza FBK

Georg Rehm, Hans Uszkoreit
(curatori, 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 978-3-642-30775-1 ISBN 978-3-642-30776-8 (eBook)
DOI 10.1007/978-3-642-30776-8
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012942722

© 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)



PREFAZIONE PREFACE

Questo Libro Bianco fa parte di una collana che intende promuovere la conoscenza in merito alle tecnologie del linguaggio e al loro potenziale. Si rivolge, tra gli altri, ai giornalisti, i politici, gli educatori e le comunità linguistiche. La disponibilità e l'uso delle tecnologie del linguaggio in Europa variano da lingua a lingua, e di conseguenza differiscono anche le azioni richieste per sostenere la ricerca e lo sviluppo di tali tecnologie. Gli interventi necessari dipendono da molti fattori, tra i quali la complessità di ciascuna lingua e le dimensioni della comunità che vi fa riferimento.

META-NET, una Rete di Eccellenza finanziata dalla Commissione Europea, con questa Collana di Libri Bianchi ha condotto un'analisi delle risorse e delle tecnologie linguistiche attualmente esistenti (p. 69). L'analisi si è concentrata sulle 23 lingue europee ufficiali e su altre importanti lingue nazionali e regionali d'Europa. I risultati di questa analisi indicano che per tutte le lingue considerate esistono dei deficit tecnologici enormi e significative lacune nella ricerca. L'analisi dettagliata che viene fornita, insieme a una valutazione della situazione attuale, potrà consentire di massimizzare l'impatto delle ricerche future.

A novembre 2011, META-NET è composta da 54 centri di ricerca, dislocati in 33 paesi europei (p. 65). META-NET collabora con aziende commerciali, enti governativi, industrie, organizzazioni di ricerca, compagnie produttrici di software e università europee. Insieme a queste comunità, META-NET sta creando una visione comune sulla tecnologia e un'agenda di ricerca strategica condivisa per l'Europa multilingue del 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. 69). 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 in 33 European countries (p. 65). 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.

Gli autori di questo documento sono grati agli autori del Libro Bianco sulla lingua tedesca per aver consentito di riutilizzare alcuni materiali selezionati dal loro documento [1].

Questo Libro Bianco è stato finanziato dal Settimo Programma Quadro e dal Programma di sostegno alla politica in materia di TIC (tecnologie dell'informazione e delle comunicazioni) della Commissione Europea nell'ambito dei contratti T4ME (accordo di finanziamento 249 119), CESAR (accordo di finanziamento 271 022), METANET4U (accordo di finanziamento 270 893) e META-NORD (accordo di finanziamento 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 [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).



INDICE CONTENTS

LA LINGUA ITALIANA NELL'ERA DIGITALE

1	Sommario	1
2	Le nostre lingue a rischio: Una sfida per le tecnologie del linguaggio	4
2.1	I confini linguistici frenano la società europea dell'Informazione	5
2.2	Le nostre lingue a rischio	5
2.3	La tecnologia del linguaggio è una tecnologia fondamentale	6
2.4	Le opportunità per le tecnologie linguistiche	7
2.5	Le sfide delle tecnologie linguistiche	8
2.6	L'acquisizione del linguaggio negli umani e nelle macchine	8
3	La lingua italiana nella società europea dell'informazione	10
3.1	Aspetti generali	10
3.2	Particolarità della lingua italiana	11
3.3	Sviluppi recenti	11
3.4	Iniziative per la promozione della lingua italiana	12
3.5	La lingua nel settore della formazione	12
3.6	L'italiano su Internet	13
4	Le tecnologie linguistiche per l'italiano	14
4.1	Architetture applicative	14
4.2	Ambiti applicativi principali	15
4.3	Altre aree applicative	22
4.4	Programmi formativi	24
4.5	Progetti e iniziative nazionali	24
4.6	Disponibilità di strumenti e risorse	25
4.7	Confronto fra le lingue	27
4.8	Conclusioni	28
5	META-NET	32

THE ITALIAN LANGUAGE IN THE DIGITAL AGE

1	Executive Summary	33
2	Languages at Risk: a Challenge for Language Technology	36
2.1	Language Borders Hold back the European Information Society	37
2.2	Our Languages at Risk	37
2.3	Language Technology is a Key Enabling Technology	37
2.4	Opportunities for Language Technology	38
2.5	Challenges Facing Language Technology	39
2.6	Language Acquisition in Humans and Machines	39
3	The Italian Language in the European Information Society	41
3.1	General Facts	41
3.2	Particularities of the Italian Language	41
3.3	Recent Developments	42
3.4	Official Language Protection in Italy	43
3.5	Language in Education	43
3.6	Italian on the Internet	43
4	Language Technology Support for Italian	45
4.1	Application Architectures	45
4.2	Core Application Areas	46
4.3	Other Application Areas	52
4.4	Educational Programmes	54
4.5	National Projects and Initiatives	54
4.6	Availability of Tools and Resources	55
4.7	Cross-language comparison	57
4.8	Conclusions	57
5	About META-NET	61
A	Riferimenti bibliografici – References	63
B	Membri di META-NET – META-NET Members	65
C	La Collana Libri Bianchi META-NET – The META-NET White Paper Series	69

SOMMARIO

Nel corso degli ultimi 60 anni, l'Europa è diventata una struttura politica ed economica distinta, che si caratterizza per la ricchezza e la varietà del suo patrimonio culturale e linguistico. Ciò significa che dal portoghese al polacco e dall'italiano all'islandese, la comunicazione quotidiana tra cittadini europei, così come la comunicazione nella sfera degli affari e della politica, sono inevitabilmente ostacolate da barriere linguistiche. Le istituzioni dell'UE spendono circa un miliardo di euro l'anno per mantenere la loro politica di multilinguismo, che consiste nella traduzione di testi scritti e nell'interpretariato di comunicazioni orali. Secondo alcune stime, il mercato europeo per la traduzione, l'interpretariato, la localizzazione del software e la globalizzazione dei siti web si aggira intorno a 8.4 miliardi di euro e ci si aspetta che aumenti del 10% all'anno. Ma si tratta di una spesa davvero necessaria? Nonostante questo impegno economico, i testi tradotti rappresentano solo una parte dell'informazione a disposizione della popolazione in paesi dove c'è una sola lingua predominante, come gli Stati Uniti, la Cina o il Giappone. Le moderne tecnologie del linguaggio e la ricerca linguistica possono dare un contributo significativo per abbattere questi confini linguistici. Se combinate con dispositivi e applicazioni intelligenti, le tecnologie del linguaggio in futuro saranno in grado di aiutare i cittadini europei a comunicare e fare affari facilmente tra loro anche se non parlano una lingua comune.

L'economia italiana trae vantaggio dal mercato unico europeo ma le barriere linguistiche possono portare ad una limitazione degli scambi, soprattutto per le PMI che non hanno i mezzi finanziari per invertire la situazione.

L'unica (impensabile) alternativa a questo tipo di Europa multilingue sarebbe quella di permettere a una singola lingua di acquisire una posizione dominante e finire per sostituire tutte le altre lingue.

Le tecnologie del linguaggio costruiscono ponti per il futuro dell'Europa.

Il modo più naturale per superare le barriere linguistiche sarebbe certamente quello di imparare le lingue straniere. Eppure, considerando la quantità delle lingue d'Europa – circa ottanta, tra lingue ufficiali e non – l'apprendimento delle lingue non basta da solo per le necessità della comunicazione, del commercio e del trasferimento dell'informazione tra tutti i confini linguistici. Senza il supporto della tecnologia, per esempio la traduzione automatica, la diversità linguistica dell'Europa rischia di rappresentare un ostacolo insormontabile per i cittadini europei e per l'economia, il dibattito politico e il progresso scientifico.

Le tecnologie del linguaggio hanno un ruolo chiave per fornire una soluzione sostenibile, economica e socialmente vantaggiosa al problema creato dalle barriere linguistiche.

Queste tecnologie offriranno agli attori europei enormi vantaggi, non solo all'interno del mercato comune europeo, ma anche nelle relazioni commerciali con i paesi terzi, in particolare le economie emergenti. Le soluzioni proposte dalle tecnologie del linguaggio finiranno per rappresentare un unico ponte tra le lingue d'Europa. Per raggiungere questo obiettivo e preservare la diversità