

THE SWEDISH LANGUAGE IN  
THE DIGITAL AGE

SVENSKA  
SPRÅKET I DEN  
DIGITALA  
TIDSÅLDERN

Lars Borin  
Martha D. Brandt  
Jens Edlund  
Jonas Lindh  
Mikael Parkvall



---

White Paper Series

Vitböcker

THE SWEDISH  
LANGUAGE IN  
THE DIGITAL  
AGE

SVENSKA  
SPRÅKET I DEN  
DIGITALA  
TIDSÅLDERN

Lars Borin Språkbanken, Göteborgs univ.

Martha D. Brandt Språkbanken, Göteborgs univ.

Jens Edlund Kungliga Tekniska högskolan

Jonas Lindh Språkbanken, Göteborgs univ.

Mikael Parkvall Stockholms universitet

---

Georg Rehm, Hans Uszkoreit  
(utgivare, 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 978-3-642-30831-4            ISBN 978-3-642-30832-1 (eBook)  
DOI 10.1007/978-3-642-30832-1  
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012940570

© 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)



# FÖRORD

## PREFACE

Denna vitbok ingår i en serie med information om språkteknologi och de möjligheter denna teknologi öppnar. Vitboken riktar sig till journalister, beslutsfattare, språkgemenskaper, utbildare och andra. Tillgången till och användningen av språkteknologi varierar stort mellan Europas språk. Därför krävs olika åtgärder som beror på många faktorer, t. ex. hur komplext språket är och hur stor språkgemenskap det handlar om.

META-NET, ett EU-finansierat spetsforskningsnätverk, har inventerat och analyserat tillgången till språkresurser och språkteknologi i denna vitboksserie (se s. 79). Analysen omfattar de 23 officiella EU-språken, samt ett antal andra viktiga national- och regionalspråk i Europa. Resultaten av analysen visar på avsevärda brister i teknikstöd och stort behov av forskningsinsatser överlag. Den detaljerade expertanalys och lägesbedömning som föreligger här kan förhoppningsvis bidra till att maximera framtida forskningsinsatsers effektivitet. META-NET består av 54 forskningscentra i 33 länder (i november 2011, se s. 75) som samverkar med intressenter från näringsliv (mjukvaru- och teknologiföretag, användare), offentlig sektor, ideella organisationer, språkgemenskaper och europeiska universitet. I samarbete med dessa grupper utvecklar META-NET en gemensam teknologivision och strategisk forskningsagenda för ett flerspråkigt Europa 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 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. 79). 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. 75). 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.

Författarna vill uttrycka sin tacksamhet till den tyska vitbokens författare som givit sitt tillstånd till användning av valda delar av deras text [1].

Arbetet med denna vitbok har utförts med finansiering från EU:s sjunde ramprogram och ICT PSP, inom projekten T4ME (avtal 249 119), CESAR (avtal 271 022), META-NET4U (avtal 270 893) och META-NORD (avtal 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).



# INNEHÅLL CONTENTS

## SVENSKA SPRÅKET I DEN DIGITALA TIDSÅLDERN

<b>1</b>	<b>Sammanfattning</b>	<b>1</b>
<b>2</b>	<b>Hotet mot våra språk: en utmaning för språkteknologin</b>	<b>4</b>
2.1	Språkgränser håller tillbaka det europeiska informationssamhället . . . . .	5
2.2	Hotet mot våra språk . . . . .	5
2.3	Språkteknologi är en nyckelteknologi . . . . .	6
2.4	Språkteknologins möjligheter . . . . .	6
2.5	Språkteknologins utmaningar . . . . .	7
2.6	Hur människor och maskiner lär sig språk . . . . .	7
<b>3</b>	<b>Svenska i det europeiska informationssamhället</b>	<b>9</b>
3.1	Bakgrundsfakta . . . . .	9
3.2	Karaktäristika för svenskan . . . . .	11
3.3	Utvecklingen under senare år . . . . .	11
3.4	Officiellt stöd för Sveriges språk . . . . .	12
3.5	Språk i utbildningssystemet . . . . .	12
3.6	Internationella aspekter . . . . .	13
3.7	Svenska på internet . . . . .	14
<b>4</b>	<b>Språkteknologi för svenska</b>	<b>16</b>
4.1	Tillämpnings- arkitekturer . . . . .	16
4.2	Centrala användningsområden . . . . .	17
4.3	Andra användningsområden . . . . .	25
4.4	Utbildning i språkteknologi . . . . .	27
4.5	Nationella projekt och initiativ . . . . .	28
4.6	Verktyg och resurser för svenska . . . . .	29
4.7	Tvärspråklig jämförelse . . . . .	30
4.8	Slutsatser . . . . .	31
<b>5</b>	<b>Vad är META-NET?</b>	<b>35</b>

# THE SWEDISH LANGUAGE IN THE DIGITAL AGE

<b>1</b>	<b>Executive Summary</b>	<b>37</b>
<b>2</b>	<b>Languages at Risk: a Challenge for Language Technology</b>	<b>40</b>
2.1	Language Borders Hold back the European Information Society . . . . .	41
2.2	Our Languages at Risk . . . . .	41
2.3	Language Technology is a Key Enabling Technology . . . . .	41
2.4	Opportunities for Language Technology . . . . .	42
2.5	Challenges Facing Language Technology . . . . .	43
2.6	Language Acquisition in Humans and Machines . . . . .	43
<b>3</b>	<b>The Swedish Language in the European Information Society</b>	<b>45</b>
3.1	General Facts . . . . .	45
3.2	Particularities of the Swedish Language . . . . .	47
3.3	Recent Developments . . . . .	47
3.4	Official Language Protection in Sweden . . . . .	48
3.5	Language in Education . . . . .	49
3.6	International Aspects . . . . .	49
3.7	Swedish on the internet . . . . .	51
<b>4</b>	<b>Language Technology Support for Swedish</b>	<b>52</b>
4.1	Application Architectures . . . . .	52
4.2	Core Application Areas . . . . .	53
4.3	Other Application Areas . . . . .	60
4.4	Educational Programmes . . . . .	62
4.5	National Projects and Initiatives . . . . .	62
4.6	Availability of Tools and Resources . . . . .	64
4.7	Cross-language comparison . . . . .	65
4.8	Conclusions . . . . .	66
<b>5</b>	<b>About META-NET</b>	<b>70</b>
<b>A</b>	<b>Litteratur – References</b>	<b>71</b>
<b>B</b>	<b>Medlemmar i META-NET – META-NET Members</b>	<b>75</b>
<b>C</b>	<b>META-NETs vitböcker – The META-NET White Paper Series</b>	<b>79</b>

## SAMMANFATTNING

Informationsteknologin förändrar vår vardag. Vi använder nu normalt datorn när vi skriver och redigerar text, när vi räknar, när vi söker kunskap och i allt högre grad när vi läser, lyssnar på musik, tittar på foton och filmer. Vi har en liten dator i fickan som vi använder för att ringa, skriva epost, hämta information och för underhållning, oavsett var vi är. Hur påverkas vårt språk av denna massiva digitalisering av information, kunskap och vardagskommunikation? Kommer vårt språk att förändras eller till och med försvinna?

Våra datorer är hopkopplade i ett alltmer vittförgrenat globalt nätverk. När europeer diskuterar reaktorhaveriet i Fukushima och hur det kan påverka Europas energipolitik i diskussionsföreläsningar och chattrum på nätet, handlar det i själva verket om ett antal separata diskussioner på en rad olika språk. Även om internet sammanbinder oss fysiskt, skiljer språken oss åt på samma sätt som alltid hittills. Kommer den situationen att bestå?

Många av världens 7 000 språk kommer inte att överleva i det globala informationssamhälle som vi nu i ilfart är på väg in i. Språkforskare har uppskattat att åtminstone 2 000 språk kommer att dö ut under de närmaste decennierna. Andra språk kommer att överleva i hemmen och lokala miljöer, men inte användas i större sammanhang, t. ex. i handel eller undervisning och forskning. Vilka är svenskans chanser att överleva?

Med sina 10 miljoner talare har svenskan en relativt stark position jämfört med många andra språk. Det finns ett antal public service-tevekanaler som sänder på svenska (sju i Sverige och en i Finland) samt några kommersiella kanaler. Trots att dess snara undergång ofta

har förutspåtts, är bok- och tidningsmarknaden faktiskt tämligen stabil och aktiv, och den årliga bokmässan i Göteborg är störst i sitt slag i Norden, med över 100 000 besökare.

Det har länge varit självklart att använda svenska för kommunikation i Norden, särskilt med de närbesläktade nordiska språken norska och danska. De tre språken har sammanlagt ca 20 miljoner talare, och de blandvarianter som ofta används i dessa sammanhang brukar kallas "skandinaviska". Svenska är det ena av Finlands två officiella språk och danska är skolämne på Island, Färöarna och Grönland. Nu tar engelskan dock alltmer över rollen som kommunikationsmedel över nationsgränserna i Norden, särskilt bland yngre talare och särskilt utanför Danmark, Norge och Sverige, där skandinaviska fortfarande håller ställningarna gentemot engelskan.

Klagomålen duggar tätt om den ökande användningen av engelska ord och uttryck i svenska och somliga är till och med rädda för att svenskan ska bli ett slags blandspråk. Inget tyder dock på att dessa farhågor har någon grund. Svenskan har överlevt ett massivt inflöde av nya ord och termer från tyska under medeltiden, liksom från franska under 1700-talet och början av 1800-talet. En bra motåtgärd mot hotet att förlora våra kära svenska ord och uttryck är att faktiskt använda dem – ofta och medvetet. Här brukar varken klagomål över främmande inflytande eller försök till officiell reglering av språkbruket åstadkomma särskilt mycket. Vi borde inte oro oss så mycket över att engelskan ska ta över vårt språk. Ett större hot är att det kan bli helt obrukbart



i stora delar av vår vardag. Då tänker vi inte på området som forskning, flygtrafik eller den globala penningmarknaden, där världen faktiskt behöver ett globalt *lingua franca*. Vi tänker på de många sammanhang där det centrala är nå landets medborgare, inte att kommunicera internationellt – t. ex. inrikespolitik, myndighetsväsen, administration, lagstiftning, kultur och handel.

Ett språks status beror inte bara på hur många som talar det eller hur många böcker, filmer och tevekanaler som använder det, utan även på hur väl det är representerat i digitala medier och datorprogram. Även i det avseendet ligger svenskan ganska bra till: de flesta allmänt använda internationella datorprogrammen finns i svenska versioner och den svenska Wikipedia ligger världselva i antal artiklar, precis före den kinesiska.

När det gäller språkteknologi, finns ett gott utbud av produkter, teknologier och resurser för svenska. Det finns tillämpningar och verktyg för talsyntes, taligenkänning, stavnings- och grammatikkontroll. Det finns även en rad tillämpningar för automatisk översättning som inkluderar svenska som ett av språken, även om många av dessa tillämpningar kommer till korta när det gäller att producera språkligt korrekta och idiomatiska översättningar, särskilt om svenska är målspråket. Detta beror till en del på specifika drag hos svenska språket.

Informations- och kommunikationsteknologierna står nu inför sin nästa revolution. Efter persondatorer, nätverk, miniaturisering, multimedia, mobila teknologier och molnet kommer nu en ny generation teknologier med mjukvara som erbjuder användarna en ännu bättre interaktion genom att den talar och förstår deras språk. Vi ser embryot till den utvecklingen i sådana tillämpningar som Googles fria översättningstjänst som översätter mellan 57 språk, IBM:s superdator Watson som besegrade USA-mästaren i Jeopardy och Apples mobila assistent Siri för iPhone som förstår talade kommandon och svarar på frågor på engelska, tyska, franska och japanska.

Nästa generations informationsteknologi kommer att hantera mänskligt språk till den grad att användarna kommer att kunna kommunicera på sitt eget språk med teknologin. Genom ett enkelt talgränssnitt kommer vi att kunna få våra apparater att leta fram de viktigaste nyheterna och den relevantaste informationen från världens digitala kunskapsbanker. Språkteknologi kommer att översätta automatiskt eller ge tolkningsstöd, sammanfatta samtal och dokument samt erbjuda stöd för lärande. Språkteknologi kommer t. ex. att kunna hjälpa invandrare att lära sig svenska och därmed hjälpa dem att integreras djupare i landets kultur.

Med nästa generations informations- och kommunikationsteknologier kommer vi att få se robotar i industrin och servicefunktioner, som förstår muntliga instruktioner från sina användare och utför dem, samt rapporterar i tal vad de har gjort.

För att åstadkomma detta krävs mjukvara som går långt bortom dagens enkla ordlistor, stavningskontrollprogram och uttalsregler. Teknologin måste gå vidare från enkla, fragmenterade approacher och ta ett helhetsgrepp på modelleringen av språket, där både syntax och semantik används för att förstå innebörden i frågor och för att kunna producera välformulerade och relevanta svar.

Men om vi jämför med vad som går att göra för engelska, ser vi att teknologin för svenska ligger långt efter och att avståndet just nu ökar. Efter en intensiv och framgångsrik satsning under 1980- och i synnerhet 1990-talet, har Sverige nu prioriterat ned forskning och utveckling inom språkteknologi, eftersom det finns andra nya, framväxande områden som uppfattas som mer angelägna att stödja. Därför har Sverige (och Europa i allmänhet) förlorat ett antal mycket lovande högteknologiska innovationer till USA, där forskningsstrategierna har präglats av större kontinuitet och där det har funnits bättre finansiellt stöd för kommersialisering av nya teknologier. När det handlar om teknologiinnovation, räc-