

THE ESTONIAN  
LANGUAGE IN  
THE DIGITAL  
AGE

EESTI  
KEEL  
DIGIAJASTUL

Krista Liin  
Kadri Muischnek  
Kaili Müürisep  
Kadri Vider



---

White Paper Series

Valge raamatu sari

THE ESTONIAN  
LANGUAGE IN  
THE DIGITAL  
AGE

EESTI  
KEEL  
DIGIAJASTUL

Krista Liin Tartu Ülikool

Kadri Muischnek Tartu Ülikool

Kaili Müürisep Tartu Ülikool

Kadri Vider Tartu Ülikool

---

Georg Rehm, Hans Uszkoreit  
(toimetajad, 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 978-3-642-30784-3            ISBN 978-3-642-30785-0 (eBook)  
DOI 10.1007/978-3-642-30785-0  
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012942725

© 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)



# EESSÕNA

## PREFACE

Eesti keele raport kuulub META-NETi väljaannete sarja, mille eesmärgiks on tutvustada keeletehnoloogia-alaseid teadmisi ja selle ala potentsiaali. Väljaande sihtgrupiks on õpetajad, ajakirjanikud, poliitikud, kogu keelekogukond ja teised teemast huvitatud.

Keeletehnoloogia kättesaadavus ja kasutamine on Euroopa keeliti väga erinev. Nii on ka meetmed, mida on vaja rakendada keeletehnoloogia arendamise ja uurimise edasiseks toetamiseks, erinevatele keeltele väga erinevad, sõltudes näiteks keele keerukusest ja selle kõnelejate arvust.

Euroopa Komisjoni rahastatud tippteadmiste võrgustik META-NET viis läbi keeleressursside ja -tehnoloogiate alase uurimise, mis keskendus 23 ametlikule Euroopa keelele ja ka teistele olulistele regionaalsetele keeltele Euroopas (vt lk 75). Analüüsi tulemus näitas, et kõigi keelte tehnoloogiates leidub märkimisväärsed puudujääke. Täpne ekspertanalüüs ja olukorra hindamine aitavad panustada edasise uurimistöö mõju suurendamise ja vähendada riske.

META-NET koosneb 33 riigi 54 uurimiskeskusest (vt lk 71), mis teevad koostööd tööstuse, valitsusasutuste, ülikoolide ja uurimisasutuste esindajatega. Koostöö tulemusena valmib ühine tehnoloogiline visioon, mis osana strateegilisest uurimiskavast näitab, kuidas keeletehnoloogilised rakendused saavad katta praegused uurimistöö puudujäägid aastaks 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 January 2012, 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, 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.

Selle dokumendi autorid tänavad saksa keele valge raamatu autoreid loa eest kasutada nende väljaandes sisaldunud keelest sõltumatuid materjale [1].

Selle keeleraporti koostamist rahastas 7. raamprogramm ja Euroopa Komisjoni IKT poliitika toetusprogramm lepingute T4ME (toetusleping 249 119), CESAR (toetusleping 271 022), METANET4U (toetusleping 270 893) ja META-NORD (toetusleping 270 899) kaudu.

---

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).



# SISUKORD CONTENTS

## EESTI KEEL DIGIAJASTUL

<b>1</b>	<b>Kokkuvõte</b>	<b>1</b>
<b>2</b>	<b>Oht meie keeltele ja väljakutse keeletehnoloogiale</b>	<b>4</b>
2.1	Keelepiirid tõkestavad Euroopa infoühiskonda . . . . .	5
2.2	Meie keeled on ohus . . . . .	5
2.3	Keeletehnoloogia on võtmetehnoloogia . . . . .	5
2.4	Keeletehnoloogia võimalused . . . . .	6
2.5	Keeletehnoloogia väljakutsed . . . . .	7
2.6	Kuidas inimesed ja masinad keelt omandavad . . . . .	7
<b>3</b>	<b>Eesti keel Euroopa infoühiskonnas</b>	<b>9</b>
3.1	Üldinfo . . . . .	9
3.2	Eesti keele eripärad . . . . .	9
3.3	Viimase aja arengud . . . . .	10
3.4	Keelehoole Eestis . . . . .	11
3.5	Keel ja haridus . . . . .	12
3.6	Rahvusvahelised aspektid . . . . .	12
3.7	Eesti keel internetis . . . . .	12
<b>4</b>	<b>Eesti keele keeletehnoloogiline tugi</b>	<b>14</b>
4.1	Rakenduste arhitektuur . . . . .	14
4.2	Kesksed rakendused . . . . .	15
4.3	Muud rakendusala . . . . .	22
4.4	Haridusprogrammid . . . . .	24
4.5	Riiklikud programmid ja algatused . . . . .	24
4.6	Vahendite ja ressursside kättesaadavus . . . . .	25
4.7	Keeltevaheline võrdlus . . . . .	26
4.8	Järeldused . . . . .	28
<b>5</b>	<b>META-NETist</b>	<b>31</b>

# THE ESTONIAN LANGUAGE IN THE DIGITAL AGE

<b>1</b>	<b>Executive Summary</b>	<b>33</b>
<b>2</b>	<b>Languages at Risk: a Challenge for Language Technology</b>	<b>36</b>
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 . . . . .	38
2.4	Opportunities for Language Technology . . . . .	38
2.5	Challenges Facing Language Technology . . . . .	39
2.6	Language Acquisition in Humans and Machines . . . . .	39
<b>3</b>	<b>The Estonian Language in the European Information Society</b>	<b>41</b>
3.1	General Facts . . . . .	41
3.2	Particularities of the Estonian Language . . . . .	41
3.3	Recent Developments . . . . .	43
3.4	Language Cultivation in Estonia . . . . .	43
3.5	Language in Education . . . . .	44
3.6	International Aspects . . . . .	45
3.7	Estonian on the Internet . . . . .	45
<b>4</b>	<b>Language Technology Support for Estonian</b>	<b>47</b>
4.1	Application Architectures . . . . .	47
4.2	Core Application Areas . . . . .	48
4.3	Other Application Areas . . . . .	56
4.4	Educational Programmes . . . . .	57
4.5	National Programmes and Initiatives . . . . .	58
4.6	Availability of Tools and Resources . . . . .	59
4.7	Cross-language comparison . . . . .	60
4.8	Conclusions . . . . .	61
<b>5</b>	<b>About META-NET</b>	<b>65</b>
<b>A</b>	<b>Kirjandus – References</b>	<b>67</b>
<b>B</b>	<b>META-NETi liikmed – META-NET Members</b>	<b>71</b>
<b>C</b>	<b>META-NETi Valge raamatu sari – The META-NET White Paper Series</b>	<b>75</b>

# KOKKUVÕTE

Viimase 60 aasta jooksul on Euroopas välja kujunenud küll ühtne poliitiline ja majanduslik struktuur, kuid kultuuri ja keelte osas on mitmekesisus säilinud. Keeleliised takistused pärsivad nii Euroopa kodanike omavaheolist kui ka äri- ja poliitikaringkondade suhtlust erinevates keeltes - portugali keelest poola keeleni ja kreeka keelest keldi keeleni. Euroopa Liidu asutused kulutavad aastas miljoneid eurosid mitmekeelsuspoliitika tagamiseks, s.t tõlgitakse tekste ja suulisi vestlusi. Aga kas meil oleks võimalik neid kulutusi vältida? Tänapäeva keeletehnoloogia ja keeleteadus annavad suure panuse keelebarjääri lõhkumiseks. Tulevikus aitab keeletehnoloogia koos nutikate seadmete ja programmidega eurooplastel üksteisega suhelda ja äri ajada isegi siis, kui nad ei räägi sama keelt.

---

## Keeletehnoloogia ehitab sillad Euroopa tulevikku.

---

Üks võimalus (kuid seejuures mõeldamatu võimalus) Euroopa mitmekeelsuse probleemi lahendamiseks oleks kasutusele võtta üks domineeriv keel ja sellega teised keeled asendada.

Klassikaline moodus keelebarjääri ületamiseks on võõrkeelte õppimine. Ent tehnilise toeta on majanduse, poliitvõtlaste ja teadusarenduse tarbeks kõigi Euroopa Liidu 23 ametliku liikmesriigi keele ja 60 muu Euroopa keele omandamine kodanikele ületamatu takistus.

Lahenduseks on võtmetehnoloogiate välja arendamine. Digitaalne keeletehnoloogia hõlmab kõiki kirjaliku ja suulise keele suhtluse vorme. Seega soodustab ta koostööd, äritegevust, teadmiste jagamist ning ühiskondli-

kus ja poliitilises diskussioonis osalemist, sõltumata seejuures kasutaja võimalikust keelebarjäärist ja arvutikasutamise oskuse tasemest. Sageli on keeletehnoloogia juba keerulistesse süsteemidesse lõimitud. Tulevikus võiks keeletehnoloogilistest lahendustest moodustuda ainulaadne Euroopa keelte vaheline sild.

Eesmärgi saavutamiseks ja samas Euroopa kultuurilise ja keelelise mitmekesisuse säilitamiseks tuleb esmalt süstemaatiliselt analüüsida iga Euroopa keele lingvistilist eripära ja seda toetava keeletehnoloogia hetkeseisu.

Eesti keelt kõneleb emakeelena umbes miljon inimest ja see on Eesti Vabariigi ainuke ametlik keel. Eesti keele igapäevast kasutust reguleerib keeleseadus. Samas on Eesti tuntud e-valitsuse ja e-riigi poliitika poolest. Eesti keel teaduse ja kõrghariduse keelena tugineb pikaajalisele eestikeelse kõrghariduse ja teadustöö traditsioonile. Erinevalt enamusest Euroopa keeltest ei kuulu eesti keel indoeuroopa keelkonda. Eesti keele eripäradeks võib lugeda täishäälikute rohkust, täis- ja kaashäälikute kolme pikkust, artiklite ja grammatilise soo puudumist. Samuti on eesti keelele iseloomulik rikkalik muutemorfoloogia. Eesti keele liitsõnamoodustus on vaba ja produktiivne. Sõnajärg lauses on küllaltki vaba.

---

## Keeletehnoloogia kui võti tulevikku.

---

Praegu turul kättesaadavad automaattõlke- ja kõnetöötlusvahendid selle eesmärgini veel ei küündi. Põhilised turul tegutsejad on kasumi saamisele suunatud Põhja-Ameerika eraettevõtted. 1970ndatel hakati Euroopa Liidus tähtsustama keeletehnoloogiat kui Eu-