

DHN 2019

Digital Humanities in the Nordic Countries

Proceedings of the Digital Humanities in the Nordic Countries 4th Conference

Copenhagen, Denmark, March 5-8, 2019.

Edited by

Costanza Navarretta

Manex Agirrezabal

Bente Maegaard

DHN 2019

Digital Humanities in the Nordic Countries

Proceedings of the Digital Humanities in the Nordic Countries 4th
Conference

Copenhagen, Denmark, March 5-8, 2019.

Edited by

Costanza Navarretta

Manex Agirrezabal

Bente Maegaard

University of Copenhagen, Faculty of Humanities, Copenhagen, Denmark

Table of Contents

Preface

- [The fourth Digital Humanities Conference, DHN2019](#) 1-8
Costanza Navarretta, Manex Agirrezabal, Bente Maegaard

Papers

- [Exploring the Quality of the Digital Historical Newspaper Archive KubHist](#) 9-17
, Yvonne Adesam, Dana Dannélls, Nina Tahmasebi
- [Creating vocabulary exercises through NLP](#) 18-32
Manex Agirrezabal, Begoña Altuna, Lara Gil-Vallejo, Josu Goikoetxea, Itziar Gonzalez-Dios
- [Annotation of subtitle paraphrases using a new web tool](#) 33-48
Mikko Aulamo, Mathias Creutz, Eetu Sjöblom
- [Managing Uncertainties: Small-scale Crowdsourcing of Author Letters](#) 49-57
Jenny Bergenmar
- [Automatic Dating of Medieval Charters from Denmark](#) 58-72
Sidsel Boldsen, Patrizia Paggio
- [New applications of gaze tracking in speech science](#) 73-78
Mattias Bystedt, Jens Edlund
- [A Corpus of Regional American Language from YouTube](#) 79-91
Steven Coats
- [Faces, Fights, and Families: Topic Modeling and Gendered Themes in Two Corpora of Swedish Prose Fiction](#) 92-111
Mats Dahllöf, Karl Berglund
- [Evaluation and refinement of an enhanced OCR process for mass digitisation](#) 112-123
Dana Dannélls, Torsten Johansson, Lars Björk
- [Distinguishing Narration and Speech in Prose Fiction Dialogues](#) 124-132
Adam Ek, Mats Wirén
- [Towards fast browsing of found audio data: 11 presidents](#) 133-142
Per Fallgren, Zofia Malisz, Jens Edlund
- [Investigating visual prosody using articulography](#) 143-148
Johan Frid, Malin Svensson Lundmark, Gilbert Ambrazaitis, Susanne Schötz, David House
- [Digitization of Konkani Texts, and their Transliteration: An Initiative towards Preservation of a Language Culture](#) 149-160
Palia Tukaram Gaonkar, Andre Rafael Fernandes
- [A Humorous View into the Past: The Old Jokes Archive](#) 161-165
Mark M. Hall, Bob Nicholson
- [Towards the Automatic Classification of Speech Subjects in the Danish Parliament Corpus](#) 166-174
Dorte H. Hansen, Costanza Navarretta, Lene Offersgaard, Jürgen Wedekind
- [Creating an Atlas over Handwritten Script Signs](#) 175-180
Anders Hast, Lasse Mårtensson, Ekta Vats, Raphaela Heil
- [Multispectral Photogrammetry: 3D models highlighting traces of paint on ancient sculptures](#) 181-189

Digitization of Konkani Texts, and their Transliteration: An Initiative towards Preservation of a Language Culture

Ms. Palia Tukaram Gaonkar¹ and Dr. Andre Rafael Fernandes²

¹ Doctoral Research Scholar, Department of English, Goa University, Goa, India

eng.palia@unigoa.ac.in

² Associate Professor, Department of English, Goa University, Goa, India

rafael@unigoa.ac.in

www.unigoa.ac.in

Abstract. Konkani is the official language of the state of Goa, located on the western coast of India. This language has faced many political threats such as four hundred and fifty years of Portuguese colonization and contention with Marathi in order to be recognised as the official language, post-Liberation in 1961. It finally entered the Eighth Schedule of the Constitution of India in 1992. These hardships have diversified the nature of Konkani. It is spoken in several dialects in the state of Goa and elsewhere, and is written in five different scripts, owing to the migration of Konkani people from Goa over the centuries. There are Konkani communities in the neighbouring states of Karnataka, Kerala and Maharashtra, which are heavily influenced by the dominant local culture. Hence, Konkani is written in Devanagari, Roman, Kannada, Malayalam and Perso-Arabic scripts.

This phenomenon creates a linguistic and literary gap in the community of Konkani speakers. Persistent efforts to bridge the gap have been made, and one of them is by taking the assistance of technology. The World Konkani Centre situated in Mangalore, Karnataka, has developed a transliteration tool, *Konkanverter*, which transliterates Unicode text between four of the five writing systems of Konkani, namely Devanagari, Roman, Kannada and Malayalam. This paper reports the attempts to digitize, and transliterate an available performance play text in Konkani, from one script to another. It also explores digitization as a way of preserving Konkani texts in multi-script formats.

Keywords: Konkani, Digitization, Transliteration

1 Introduction

The official language of Goa, Konkani, also spoken in pockets of Karnataka, Kerala and Maharashtra, showed a negative growth rate according to the Census of 2011[1]. Of the total population of India, 0.19% of the population claimed it to be their mother