Computational Sanskrit & Digital Humanities

Selected papers presented
at
the 17^{th} World Sanskrit Conference

University of British Columbia, Vancouver
9–13 July 2018

D K Publishers Distributors Pvt. Ltd.

Edited by
Gérard Huet & Amba Kulkarni

First Published: 2018

ISBN: **978-93-87212-10-7**

Published by
D K Publishers Distributors Pvt. Ltd.
4224/1, Ansari Road, Daryaganj,
New Delhi-110002, India
Ph: 011-41562573-77 Fax: 011-41562578
E-mail: info@dkpd.com
www.dkpd.com

Printed at:
D K Fine Art Press P Ltd.
A-6, Nimri Colony, Ashok Vihar, Phase-IV
Delhi-110052, India

Preface

This volume contains edited versions of papers accepted for presentation at the 17th World Sanskrit Conference in July 2018 in Vancouver, Canada. A special track of the conference was reserved for the topic "Computational Sanskrit & Digital Humanities", with the intent to cover not only recent advances in each of the now mature fields of Sanskrit Computational Linguistics and Sanskrit Digital Libraries, but to encourage cooperative efforts between scholars of the two communities, and prepare the emergence of grammatically informed digital Sanskrit corpus. Due to its rather technical nature, the contributions were not judged on mere abstracts, but on submitted full papers reviewed by a Progam Committee.

We would like to thank the Program Committee of our track for their work:

- Dr Tanuja Ajotikar, Belgavi, Karnataka
- Pr Stefen Baums, University of Munich
- Pr Yigal Bronner, Hebrew University of Jerusalem
- Pr Pawan Goyal, IIT Kharagpur
- Dr Oliver Hellwig, Düsseldorf University
- Dr Gérard Huet, Inria Paris (co-chair)
- Pr Girish Nath Jha, JNU, Delhi
- Pr Amba Kulkarni, University of Hyderabad (co-chair)
- Dr Pawan Kumar, Chinmaya Vishwavidyapeeth, Veliyanad
- Pr Andrew Ollett, Harvard University
- Dr Dhaval Patel, I.A.S. Officer, Gujarat
- Pr Srinivasa Varakhedi, KSU, Bengaluru

14 contributions were accepted, revised along referees' recommendations, and finely edited to form this collection.

The first two papers concern the problem of proper mechanical simulation of Pāṇini's Aṣṭadyāyī. In "A Functional Core for the Computational

Aṣṭadyāyī", Samir Janardan Sohoni and Malhar A. Kulkarni present an original architecture for such a simulator, based on concepts from functional programming. In their model, each Pāṇinian sūtra translates as a Haskell module, an elegant effective formalisation. They explain an algorithm for sūtra-conflict assessment and resolution, discuss visibility and termination, and exhibit a formal derivation of word *bhavati* as showcase.

A different computational framework for the same problem is offered by Sarada Susarla, Tilak M. Rao and Sai Susarla in their paper "PAIAS: Pāṇini Aṣṭadyāyī Interpreter As a Service". They explain their development of a Web service usable as a Sanskrit grammatical assistant, implementing directly the Pāninian mechanisms. Here sūtras are records in a database in the JSON format, managed by a Python library. They pay particular attention to the meta-rules of the grammar, and specially to defining sūtras. Thay refrain from expanding such definitions in operative sūtras, but insist on their emulation along the grammatical processing.

These two papers conceptualise two computational models of Pāninian grammar that are strikingly different in their software architecture. However, when one examines examples of sūtra representations in both systems, the information content looks very similar, which may suggest some future inter-operability of these two interesting tools.

In the general area of mechanical analysis of Sanskrit text, we have several contributions at various levels. At the level of semantic roles analysis, at the heart of dependency parsing, Sanjeev Panchal and Amba Kulkarni present possible solutions to the complementary problem of ambiguity. In their paper "Yogyatā as an absence of non-congruity" they explain various definitions used by Sanskrit grammarians to express compatibility, and how to use these definitions to reduce ambiguity in dependency analysis.

The next paper, "An 'Ekalavya' Approach to Learning Context Free Grammar Rules for Sanskrit Using Adaptor Grammar", by Amrith Krishna, Bodhisattwa Prasad Majumder, Anil Kumar Boga, and Pawan Goyal, presents an innovative use of adaptor grammars to learn patterns in Sanskrit text definable as context-free languages. They present applications of their techniques to word reordering tasks in Sanskrit, a premiminary step towards recovering prose ordering from poetry, a crucial problem in Sanskrit.

Concerning meter recognition, we have a contribution of Shreevatsa Rajagopalan on "A user-friendly tool for metrical analysis of Sanskrit verse". The main feature of this new metrical analysis tool, available either as a

Web service or as a software library, is its robustness and its guidance in error-correction.

Two more contributions use statistical techniques (Big Data) for improving various Sanskrit-related tasks.

For instance, in the field of optical character recognition, the contribution by Devaraja Adiga, Rohit Saluja, Vaibhav Agrawal, Ganesh Ramakrishnan, Parag Chaudhuri, K. Ramasubramanian and Malhar Kulkarni on "Improving the learnability of classifiers for Sanskrit OCR corrections".

In the same vein of statistical techniques, Nikhil Chaturvedi and Rahul Garg present "A Tool for Transliteration of Bilingual Texts Involving Sanskrit", which accommodates smoothly text mixing various encodings.

While much of the work in Sanskrit computational linguistics is on Classical Sanskrit, researchers are also applying computational techniques to Vedic Sanskrit. One such effort is a detailed formalisation of Vedic recitation phonology by Balasubramanian Ramakrishnan: "Modeling the Phonology of Consonant Duplication and Allied Changes in the Recitation of Tamil Taittiryaka-s".

On a more theoretical perspective on Sanskrit syntax, Brendan Gillon presents a formalisation of Sanskrit complements in terms of the categorial grammars framework. His paper "Word complementation in Sanskrit treated by a modest generalization of categorial grammar" explains modified versions of the cancellation rules that aim at accommodating free word order. This raises the theoretical problem of the distinction between complements and modifiers in Sanskrit.

Turning to the Digital Humanities theme, we have a number of contributions. In "TEITagger: Raising the standard for digital texts to facilitate interchange with linguistic software", Peter Scharf discusses how fine-grain XML representation of corpus within the Text Encoding Initiative standard allows the inter-communication between digital Sanskrit libraries and grammatical tools such as parsers as well as meter analysis tools.

A complementary proposal is discussed in the paper "Preliminary Design of a Sanskrit Corpus Manager" by Gérard Huet and Idir Lankri. They propose a scheme for a fine-grained representation of Sanskrit corpus allowing inter-textuality phenomena such as sharing of sections of text, but also variance of readings. They propose to use grammatical analysis tools to help annotators feeding digital libraries with grammatical information using modern cooperative work software. They demonstrate a prototype of such a tool, in the framework of the Sanskrit Heritage platform.

Moving towards philological concerns such as critical editions, the paper "Enriching the digital edition of the Kāśikāvṛtti by adding variants from the Nyāsa and Padamañjarī", by Tanuja P. Ajotikar, Anuja P. Ajotikar, and Peter M. Scharf, discusses the problem of managing complex information from recensions and variants. It argues for a disciplined method of using TEI structure to represent this information in machine-manipulable ways, and demonstates its use on processing variants of the Kāśikāvṛtti, the major commentary of the Aṣṭadyāyī.

In the same area of software-aided philology, the contribution "From the web to the desktop: IIIF-Pack, a document format for manuscripts using Linked Data standards", by Timothy Bellefleur, presents a proposal for a common format fit to manage complex information about corpus recensions in various formats, including images of manuscripts. This is in view of facilitating the interchange of such data by various teams using this common format. His proposal uses state-of-the art standards of hypertext. It has already been put to use in an interactive software platform to manage recensions for the critical edition of the Vetālapañcaviṃśati by Pr. Adheesh Sathaye.

The volume concludes with the contribution "New Vistas to study Bhartṛhari: Cognitive NLP" by Jayashree Aanand Gajjam, Diptesh Kanojia, and Malhar Kulkarni which presents a highly original research on cognitive linguistics in Sanskrit, by comparing the results of experiments with eye-tracking equipment with theories of linguistic cognition by Bhartṛhari.

We thank the numerous experts who helped us in the review process, and all our authors who responded positively to the reviewer's comments and improved their manuscripts accordingly. We thank the entire 17^{th} WSC organising committee, led by Pr. Adheesh Sathaye, which provided us the necessary logistic support for the organisation of this section.

GÉRARD HUET & AMBA KULKARNI

Contributors

Devaraja Adiga
Department of Humanities and Social Sciences,
Indian Institute of Technology Bombay,
Powai, Mumbai, India.
pdadiga@iitb.ac.in

Vaibhav Agrawal
Indian Institute of Technology,
Kharagpur,
India.
vaibhav@iitkgp.ac.in

Anuja Ajotikar
Shan State Buddhist University,
Myanmar
anujaajotikar@gmail.com

Tanuja Ajotikar
KAHER's Shri B. M. Kankanwadi Ayurveda Mahavidyalaya,
Belagavi,
India.
gtanu30@gmail.com

Timothy Bellefleur
Department of Asian Studies,
University of British Columbia, Vancouver
tbelle@alumni.ubc.ca

ANIL KUMAR BOGA
Department of Computer Science and Engineering,
Indian Institute of Technology,
Kharagpur, India.
bogaanil.009@gmail.com

NIKHIL CHATURVEDI
Department of Computer Science and Engineering,
Indian Institute of Technology,
New Delhi
cs5130291@cse.iitd.ac.in

PARAG CHAUDHURI
Indian Institute of Technology Bombay,
Powai, Mumbai, India.
paragc@cse.iitb.ac.in

JAYASHREE AANAND GAJJAM
Department of Humanities and Social Sciences,
Indian Institute of Technology Bombay,
Powai, Mumbai, India.
jayashree_aanand@iitb.ac.in

RAHUL GARG
Department of Computer Science and Engineering,
Indian Institute of Technology,
New Delhi
rahulgarg@cse.iitd.ac.in

BRENDAN S. GILLON
McGill University
Montreal, Quebec
H3A 1T7 Canada
brendan.gillon@mcgill.ca

Pawan Goyal
Department of Computer Science and Engineering,
Indian Institute of Technology,
Kharagpur, India.
pawang@cse.iitkgp.ernet.in

Gérard Huet
Inria Paris Center,
France.
Gerard.Huet@inria.fr

Diptesh Kanojia
IITB-Monash Research Academy, Powai, Mumbai, India
diptesh@iitb.ac.in

Amrith Krishna
Department of Computer Science and Engineering,
Indian Institute of Technology,
Kharagpur, India.
amrith.krishna@cse.iitkgp.ernet.in

Amba Kulkarni
Department of Sanskrit Studies,
University of Hyderabad,
Hyderabad, India.
apksh@uohyd.ernet.in

Malhar Kulkarni
Department of Humanities and Social Sciences,
Indian Institute of Technology Bombay,
Powai, Mumbai, India.
malhar@hss.iitb.ac.in

Idir Lankri
Université Paris Diderot,
Paris
lankri.idir@gmail.com

Bodhisattwa Prasad Majumder
Walmart Labs
India.
bodhisattwapm2017@email.iimcal.ac.in

Sanjeev Panchal
Department of Sanskrit Studies,
University of Hyderabad,
Hyderabad, India.
snjvpnchl@gmail.com

Shreevatsa Rajagopalan
Independent Scholar
1035 Aster Ave 1107
Sunnyvale, CA 94086, USA
shreevatsa.public@gmail.com

Balasubramanian Ramakrishnan
Independent Scholar
145 Littleton Rd
Harvard, MA 01451, USA
balasr@acm.org

Ganesh Ramakrishnan
Indian Institute of Technology Bombay,
Powai, Mumbai, India.
ganesh@cse.iitb.ac.in

K. Ramasubramanian
Department of Humanities and Social Sciences,
Indian Institute of Technology Bombay,
Powai, Mumbai, India.
ram@hss.iitb.ac.in

Rohit Saluja
IITB-Monash Research Academy,
Powai, Mumbai, India.
rohitsaluja@cse.iitb.ac.in

Peter Scharf
The Sanskrit Library,
Providence, Rhode Island, U.S.A.
and
Language Technologies Research Center,
Indian Institute of Information Technology,
Hyderabad, India.
scharf@sanskritlibrary.org

Samir Janardan Sohoni
Department of Humanities and Social Sciences,
Indian Institute of Technology Bombay,
Powai, Mumbai, India.
sohoni@hotmail.com

Contents