

5th Workshop on South and Southeast Asian NLP

(WSSANLP 2014+)

Held at COLING 2014

**Dublin, Ireland
23 August 2014**

ISBN: 978-1-63439-223-5

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2014) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2014)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Towards Identifying Hindi/Urdu Noun Templates in Support of a Large-Scale LFG Grammar</i> Sebastian Sulger and Ashwini Vaidya	1
<i>Konkanverter - A Finite State Transducer based Statistical Machine Transliteration Engine for Konkani Language</i> Vinodh Rajan	11
<i>Integrating Dictionaries into an Unsupervised Model for Myanmar Word Segmentation</i> Ye Kyaw Thu, Andrew Finch, Eichiro SUMITA and Yoshinori Sagisaka	20
<i>A Framework for Learning Morphology using Suffix Association Matrix</i> Shilpa Desai, Jyoti Pawar and Pushpak Bhattacharyya	28
<i>English to Urdu Statistical Machine Translation: Establishing a Baseline</i> Bushra Jawaid, Amir Kamran and Ondrej Bojar	37
<i>A hybrid approach for automatic clause boundary identification in Hindi</i> Rahul Sharma and Soma Paul	43
<i>RBMT as an alternative to SMT for under-resourced languages</i> Guillaume de Malézieux, Amélie Bosc and Vincent Berment	50
<i>Developing an interlingual translation lexicon using WordNets and Grammatical Framework</i> Shafqat Mumtaz Virk, K.V.S Prasad, Aarne Ranta and Krasimir Angelov	55
<i>A Dictionary Data Processing Environment and Its Application in Algorithmic Processing of Pali Dictionary Data for Future NLP Tasks</i> Jürgen Knauth and David Alfter	65
<i>Constituent structure representation of Pashto Endoclitics</i> Azizud Din, Bali Ranaivo-Malançon and M. G. Abbas Malik	74
<i>Real Time Early-stage Influenza Detection with Emotion Factors from Sina Microblog</i> Xiao Sun, Jiaqi Ye and Fuji Ren	80
<i>Building English-Vietnamese Named Entity Corpus with Aligned Bilingual News Articles</i> Quoc Hung Ngo, Dinh Dien and Werner Winiwarter	85
<i>Character-Cluster-Based Segmentation using Monolingual and Bilingual Information for Statistical Machine Translation</i> Vipas Sutantayawalee, Peerachet Porkeaw, Thepchai Supnithi, Prachya Boonkwan and Sitthaa Phaholphinyo	94
<i>A rule based approach for automatic clause boundary detection and classification in Hindi</i> Rahul Sharma	102