37th Pacific Asia Conference on Language, Information and Computation (PACLIC 37)

Hong Kong, China
2-4 December 2023

ISBN: 978-1-7138-8992-2

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© (2023) by PACLIC 37 Organizing Committee and PACLIC Steering Committee
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

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: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

Sensitivity and Robustness of Large Language Models to Prompt Template in Japanese Text Classification Tasks 1
Chengguang Gan, Tatsunori Mori

Large Scale Evaluation of End-To-End Pipeline of Speaker to Dialogue Attribution in Japanese Novels........ 12
Yuki Zenimoto, Shinzan Komata, Takehito Utsuro

Machine Reading Comprehension for Vietnamese Customer Reviews: Task, Corpus and Baseline Models 24
Tinh Pham Phuc Do, Ngoc Dinh Duy Cao, Nhan Thanh Nguyen, Tin Van Huynh, Kiet Van Nguyen

Simultaneous Domain Adaptation of Tokenization and Machine Translation........ 36
Taisei Enomoto, Tosho Hirasawa, Hwichan Kim, Teruaki Oka, Mamoru Komachi

UPOS-DEPREL Mismatches: Detecting Annotation Errors and Improving UD Guidelines Based on Linguistic Knowledge 46
Tsy Yih, Zheyuan Dai

LIDA: Lexical-Based Imbalanced Data Augmentation for Content Moderation 59
Guangming Huang, Yunfei Long, Cunjin Luo, Yingya Li

Data Augmentation for SentRev Using Back-Translation of Lexical Bundles........ 70
Zhendong Du, Kenji Hashimoto

The WEAVE 2.0 Corpus: Role Labelled Synthetic Chemical Procedures from Patents with Chemical Named Entities 80
Shubhangi Dutta, Manish Shrivastava, Prabhakar Bhimalapuram

Low-Frequency Long-Distance Dependencies as "Long Tails" 89
Masanori Oya

Data-Augmented Task-Oriented Dialogue Response Generation with Domain Adaptation 96
Yan Pan, Davide Cadamuro, Georg Groh

Word Familiarity Rate Estimation for Japanese Functional Words Using a Bayesian Linear Mixed Model 107
Bocheng Chen, Masayuki Asahara

What Should I Ask: A Knowledge-Driven Approach for Follow-Up Questions Generation in Conversational Surveys........113
Yubin Ge, Ziang Xiao, Jana Diesner, Heng Ji, Karrie Karahalios, Hari Sundaram

Construction of Evaluation Dataset for Japanese Lexical Semantic Change Detection 125
Zhidong Ling, Taichi Aida, Teruaki Oka, Mamoru Komachi

Spatial Information Annotation Based on the Double Cross Model........ 137
Yoshiko Kawabata, Mai Omura, Masayuki Asahara, Johane Takeuchi

Critical Thinking Skills in an Oral Reading Comprehension Test: A Curious Case of the Individuals with Tourette Syndrome 145
Aileen Bautista, Rochelle Irene Lucas

L3Cube-IndicSBERT: A Simple Approach for Learning Cross-Lingual Sentence Representations Using Multilingual BERT 154
Samruddhi Deode, Janhavi Gadre, Aditi Kajale, Ananya Joshi, Raviraj Joshi

Encoding Syllables in Zhangzhou Southern Min 164
Yishan Huang

Japanese Benefactive Passives Are Difficult to Comprehend than Benefactive Actives 174
Masataka Ogawa

Data Augmentation by Shuffling Phrases in Recognizing Textual Entailment 194
Kyosuke Takahagi, Hiroyuki Shinnou

All-Words Word Sense Disambiguation for Historical Japanese 201
Soma Asada, Kanako Komiya, Masayuki Asahara

Building a Topic-Oriented Corpus and Its Application for Language Teaching 210
Naoki Nakamata

Improving Zero-Shot Dependency Parsing by Unsupervised Learning 217
Jiannan Mao, Chenchen Ding, Hour Kaing, Hideki Tanaka, Masao Utiyama, Tadahiro Matsumoto

Directional Constructions in Taiwan Southern Min: Complex Predicates Or Compounds 227
Huei-Ling Lin

Practical Approaches for Low-Resource Named Entity Recognition of Filipino Telecommunications Domain 234
Kyle Chan, Kaye Ann De Las Alas, Charles Orcena, Dan John Velasco, Qyle John San Juan, Charibeth Cheng

Solving Label Variation in Scientific Information Extraction Via Multi-Task Learning 243
Dong Pham, Xanh Ho, Quang Thuy Ha, Akiko Aizawa

Multi-Task Learning for Emotion Recognition in Conversation with Emotion Shift 257
Juntao Wang, Tsunenori Mine

Rapid Speaker Adaptation in Low Resource Text to Speech Systems Using Synthetic Data and Transfer Learning 267
Raviraj Joshi, Nikesh Garera

L3Cube-MahaSent-MD: A Multi-Domain Marathi Sentiment Analysis Dataset and Transformer Models 274
Aabha Pingle, Aditya Vyawahare, Isha Joshi, Rahul Tangsali, Raviraj Joshi

Improving Cross-Domain Hate Speech Generalizability with Emotion Knowledge 282
Shi Yin Hong, Susan Gauch

Exploring Affordance and Situated Meaning in Image Captions: A Multimodal Analysis 293
Pin-Er Chen, Po-Ya Angela Wang, Hsin-Yu Chou, Yu-Hsiang Tseng, Shu-Kai Hsieh

From Annotation to Analysis: Exploring Conversational Dynamics with rezonateR 303
Ryan Ka Yau Lai

Information Compression Via Eliding Verb Phrase: A Dependency-Based Study 314
Zheyuan Dai, Chenliang Zhou, Haitao Liu

Discovering Phonesthemic Clusters in Readings of Kanji Characters Toward Exploring Phonestheme in Japanese 322
Akira Yoshida, Chihaya Matsuhira, Hirotaka Kato, Takatsugu Hirayama, Takahiro Komamizu, Ichiro Ide

Generating Character Lines in Four-Panel Manga 338
Michimasa Inaba

SumRec: A Framework for Recommendation Using Open-Domain Dialogue 349
Ryutaro Asahara, Masaki Takahashi, Chiho Iwahashi, Michimasa Inaba

Linguistic and Paralinguistic Features Influencing Reliability Judgments of Thai Twitter Reviews 364
Warisaraporn Limprasert, Nanthicha Angsuwichitkul, Francesco Burroni

DialoGen: Generalized Long-Range Context Representation for Dialogue Systems........ 372
Suvodip Dey, Maunendra Sankar Desarkar, Asif Ekbal, Srijith P. K.

ViASR: A Novel Benchmark Dataset and Methods for Vietnamese Automatic Speech Recognition 387
Binh Nguyen, Son Huynh, Quoc Khanh Tran, An Le Tran-Hoai, Trong An Nguyen, Nguyen Tung Doan Tran, Thuy An Phan Thi, Le Thanh Nguyen, Hieu Nghia Nguyen, Dang Huynh

A Corpus for Named Entity Recognition in Chinese Novels with Multi-Genres 398
Hanjie Zhao, Jinge Xie, Yuchen Yan, Yuxiang Jia, Yawen Ye, Hongying Zan

Contrasting the Chinese Bei-Passive and the English Be-Passive Through Dependency Distance 406
Yonghui Xie, Ruochen Niu, Haitao Liu

A Methodology for Exploring Experiences of Individuals with Tourette Syndrome in Oral and Silent Reading Assessments 416
Aileen Bautista, Macario Cordel II, Cely Magpantay, Catherine Manuela Ramos

Sinhala-English Word Embedding Alignment: Introducing Datasets and Benchmark for a Low Resource Language 424
Kasun Wickramasinghe, Nisansa De Silva

An Empirical, Corpus-Based, Approach to Cantonese Nominal Expressions........ 436
Gregoire Winterstein, David Vergnaud, Hannah Hoi Tung Yu, Jeremie Lupien, Laperle Samuel, Pei Sui Luk, Christopher Davis

Developing an Annotated Persian Dataset from COVID-19 News for Enhanced Fake News Detection 446
Foroogh Zahed, Seyedeh Fatemeh Ebrahimi, Mohammad Bahrani, Alireza Mansouri

Sibilant Production in Taiwan Mandarin: Untangling the Effects of Linguistic and Social Variables 456
Yu-Leng Lin, Jiun-Shiung Wu

Effectiveness of ChatGPT in Korean Grammatical Error Correction 464
Junghwan Maeng, Jinghang Gu, Sun-A Kim

Enhanced Retrieve-Edit-Rerank Framework with kNN-MT........ 473
Xiaotian Wang, Takuya Tamura, Takehito Utsuro, Masaaki Nagata

Partially Randomizing Transformer Weights for Dialogue Response Diversity 486
Jing Yang Lee, Kong Aik Lee, Woon-Seng Gan

Constructing a Japanese Business Email Corpus Based on Social Situations 499
Muxuan Liu, Tatsuya Ishigaki, Yusuke Miyao, Hiroya Takamura, Ichiro Kobayashi

Automatic Identification of Assistance Needs in Disaster Situations Using Hybrid Word Embedding Techniques 510
Sarra Chaiir, Malika Charrad, Narjes Bellamine Ben Saoud

Estimating Lexical Complexity in Multi-Domain Settings for the Russian Language 516
Aleksei Abramov, Vladimir Ivanov, Valery Solovyev

Conceptualizing Philippine Language Policy Using Liberal Neutrality Philosophical Model: A Policy Brief 527
Bernard Barruga

Estimating the Likelihood of Words Being Known with Corpus Analysis and K-Means Clustering Algorithm 535
Tong Zhu, Derek Irwin, Yanhui Zhang, Renjie Wu, Xiaoyi Jiang

Lyrics Generation Applying Metaphor Generation 543
Kazuki Murakami, Asuka Terai

A Follow-Up Study on Evaluation Metrics Using Follow-Up Utterances 552
Toshiki Kawamoto, Yuki Okano, Takato Yamazaki, Toshinori Sato, Kotaro Funakoshi, Manabu Okumura

The Cross-Linguistic Variations in Dependency Distance Minimization and Its Potential Explanations 559
Ruochen Niu, Yaqin Wang, Haitao Liu

A Corpus-Based Investigation of Occurrences and Functions of Frame Markers in Filipino University Students¡¯COVID-19 Written Narratives 570
Gregorio Ebron Jr., Romualdo Mabuan, Shirley Dita

Self-MI: Efficient Multimodal Fusion Via Self-Supervised Multi-Task Learning with Auxiliary Mutual Information Maximization 582
Cam-Van Nguyen Thi, Ngoc-Hoa Nguyen Thi, Duc-Trong Le, Quang-Thuy Ha

Passage-Based BM25 Hard Negatives: A Simple and Effective Negative Sampling Strategy for Dense Retrieval 591
Thanh-Do Nguyen, Chi Minh Bui, Thi-Hai-Yen Vuong, Xuan-Hieu Phan

Efficient GPT Model Pre-Training Using Tensor Train Matrix Representation 600
Viktoriia Chekalina, Georgiy Novikov, Julia Gusak, Alexander Panchenko, Ivan Oseledets

From Data to Dialogue: Leveraging the Structure of Knowledge Graphs for Conversational Exploratory Search 609
Phillip Schneider, Nils Rehtanz, Kristiina Jokinen, Florian Matthes

Mukhyansh: A Headline Generation Dataset for Indic Languages 620
Lokesh Madasu, Gopichand Kanumolu, Nirmal Surange, Manish Shrivastava

Large Language Models Meet Knowledge Graphs to Answer Factoid Questions 635
Mikhail Salnikov, Hai Le, Prateek Rajput, Irina Nikishina, Pavel Braslavski, Valentin Malykh, Alexander Panchenko

Vocabulary Replacement in SentencePiece for Domain Adaptation 645
Teruno Kajiura, Shiho Takano, Tatsuya Hiraoka, Kimio Kuramitsu

Revisiting Korean Corpus Studies Through Technological Advances 653
Won Ik Cho, Sangwhan Moon, Youngsook Song

Toward a Corpus-Based Identification of Nominal Relationality and Uniqueness: A Constructionist Approach 661
Kazuho Kambara, Taishi Chika

Automatic Transcript Generation from Presentation Slides 670
Nguyen Xuan Vu Nguyen, Quang Huy Ngo, Quang Nhat Minh Pham

Vec2Gloss: Definition Modeling Leveraging Contextualized Vectors with Wordnet Gloss........ 679
Yu-Hsiang Tseng, Mao-Chang Ku, Wei-Ling Chen, Yu-Lin Chang, Shu-Kai Hsieh

Subwords to Word Back Composition for Morphologically Rich Languages in Neural Machine Translation........ 691
Telem Joyson Singh, Sanasam Ranbir Singh, Priyankoo Sarmah

The Association of Second Language Proficiency with Nonverbal Behaviors 701
Katsunori Kotani, Takehiko Yoshimi

Paper Recommendation Using Citation Contexts in Scholarly Documents 710
Tomoki Ikoma, Shigeki Matsubara

Using Deepfake Technologies for Word Emphasis Detection 717
Eran Kaufman, Lee-Ad Gottlieb, Dina Mayzlish, Or Tiram, Hila Wiesel, Nofar Yosef

A Computational Study of Matrix Decomposition Methods for Compression of Pre-Trained Transformers........ 723
Sergey Pletenev, Viktoriia Chekalina, Daniil Moskovskiy, Mikhail Seleznev, Sergey Zagoruyko, Alexander Panchenko

Image Caption Synthesis for Low Resource Assamese Language Using Bi-LSTM with Bilinear Attention........ 743
Pankaj Choudhury, Prithwijit Guha, Sukumar Nandi

Turn Design, Resonance and Epistemic Stance in the Diamond Sutra: A Dialogic Constructionist Approach 753
Ryan Ka Yau Lai, Lily Zihe Yin, Alice Yimeng Zhang, Yuting Jiang, Bill Shiyang Xin, Junwei Gao

Computational Semantics and Evaluation Benchmark for Interrogative Sentences Via Combinatory Categorial Grammar 764
Hayate Funakura, Koji Mineshima

EDAL: Entropy Based Dynamic Attention Loss for HateSpeech Classification 775
Md Fahim, Dr. Amin Ahsan Ali, Md Ashraful Amin, Akm Mahbubur Rahman

Training Generative Question-Answering on Synthetic Data Obtained from an Instruct-Tuned Model 786
Kosuke Takahashi, Takahiro Omi, Kosuke Arima, Tatsuya Ishigaki

CebuaNER: A New Baseline Cebuano Named Entity Recognition Model 792
Ma. Beatrice Emanuela Pilar, Dane Dedoroy, Ellyza Mari Papas, Mary Loise Buenaventura, Myron Darrel Montefalcon, Jay Rhald Padilla, Joseph Marvin Imperial, Mideth Abisado, Lany Maceda

Assamese Back Transliteration - An Empirical Study Over Canonical and Non-Canonical Datasets 801
Hemanta Baruah, Sanasam Ranbir Singh, Priyankoo Sarmah

Improving Multi-Lingual Medical Term Normalization to Address the Long-Tail Problem 809
Beatrice Portelli, Simone Scaboro, Giuseppe Serra

Sparse Neural Retrieval Model for Efficient Cross-Domain Retrieval-Based Question Answering................ 819
Kosuke Nishida, Naoki Yoshinaga, Kyosuke Nishida

ViPubmedDeBERTa: A Pre-Trained Model for Vietnamese Biomedical Text 831
Manh Tran-Tien, Huu-Loi Le, Dang Nhat Minh, T. Tran Khang, Huy-The Vu, Nguyen Minh-Tien

Enhanced Transformer Architecture for Natural Language Processing.. 841
Woohyeon Moon, Taeyoung Kim, Bumgeun Park, Dongsoo Har

RealPersonaChat: A Realistic Persona Chat Corpus with Interlocutors' Own Personalities 852
Sanae Yamashita, Koji Inoue, Ao Guo, Shota Mochizuki, Tatsuya Kawahara, Ryuichiro Higashinaka

Comprehensive Evaluation of Translation Error Correction Models 862
Masatoshi Otake, Yusuke Miyao

Automatic Insertion of Commas and Linefeeds into Lecture Transcripts Based on Multi-Task Learning 872
Zhicheng Fang, Masaki Murata, Shigeki Matsubara

Counterfactual Adversarial Training for Improving Robustness of Pre-Trained Language Models 881
Hoai Linh Luu, Naoya Inoue

Fine-Grained Contract NER Using Instruction Based Mode 889
Hiranmai Adibhatla, Pavan Baswani, Manish Shrivastava

Tracing Social Change Through Metaphor: A Diachronic Corpus-Assisted Analysis.................................... 903
Winnie Huiheng Zeng, Kathleen Ahrens, Chu-Ren Huang

Existence Justifies Reason: A Data Analysis on Chinese Classifiers Based on Eye Tracking and Transformers.. 912
Yu Wang, Emmanuele Chersoni, Chu-Ren Huang

Author Index