PACLIC 2023 – Program Draft

**Day 1: Sat, Dec 2**

2:15-2:30 pm Introduction

2:30-3:15 pm Invited Talk: Dr. Enrico Santus (Bloomberg)

3:15-4:30 pm Parallel Sessions 1

1a **Psycholinguistics and Experimental Linguistics**

* Masataka Ogawa. *Japanese benefactive passives are more difficult to comprehend than benefactive actives*
* Bocheng Chen and Masayuki Asahara. *Word Familiarity Rate Estimation for Japanese Functional Words Using a Bayesian Linear Mixed Model*
* Aileen Bautista, Macario Ii Cordel, Cely Magpantay and Catherine Manuela Ramos. *A Methodology for Exploring Experiences of Individuals with Tourette Syndrome in Oral and Silent Reading Assessments*

1b **Language Models 1**

* Mikhail Salnikov, Hai Le, Prateek Rajput, Irina Nikishina, Pavel Braslavski, Valentin Malykh and Alexander Panchenko*. Large Language Models Meet Knowledge Graphs to Answer Factoid Questions*
* Hoai Linh Luu and Naoya Inoue. *Counterfactual Adversarial Training for Improving Robustness of Pre-trained Language Models*
* Sergey Pletenev, Viktoriia Chekalina, Daniil Moskovskiy, Mikhail Seleznev, Sergey Zagoruyko and Alexander Panchenko*. A computational study of matrix decomposition methods for compression of pre-trained Transformers*

4:30-4:45 pm Coffee Break

4:45-6:00 pm Parallel Sessions 2

2a **Corpus Linguistics and Discourse Analysis 1**

* Kazuho Kambara and Taishi Chika. *Toward a corpus-based identification of nominal relationality and uniqueness: A constructionist approach*
* Winnie Huiheng Zeng, Kathleen Ahrens and Chu-Ren Huang. *Tracing Social Change through Metaphor: A Diachronic Corpus-Assisted Analysis*
* Muxuan Liu, Tatsuya Ishigaki, Yusuke Miyao, Hiroya Takamura and Ichiro Kobayashi. *Constructing a Japanese Business Email Corpus Based on Social Situations*

2b **NLP for Social Media**

* Sarra Chaiir, Malika Charrad and Narjès Bellamine Ben Saoud. *Automatic Identification of Assistance Needs in Disaster Situations Using Hybrid Word Embedding Techniques*
* Foroogh Zahed, Seyedeh Fatemeh Ebrahimi, Mohammad Bahrani and Alireza Mansouri. *Developing an Annotated Persian Dataset from COVID-19 News for Enhanced Fake News Detection*
* Warisaraporn Limprasert, Nanthicha Angsuwichitkul and Francesco Burroni. *Linguistic and Paralinguistic Features influencing Reliability Judgments of Thai Twitter Reviews*

7:00 pm Welcome Reception at Dim Sum Restaurant

**Day 2: Sun, Dec 3**

9:15-10:15 am Keynote Speaker 1: Prof. Lori Levin (Carnegie Mellon University)

10:15-10:30 am Coffee Break

10:30-11:45 am Parallel Sessions 3

3a **NLP for Dialogue and Conversation 1**

* Yubin Ge, Ziang Xiao, Jana Diesner, Heng Ji, Karrie Karahalios and Hari Sundaram. *What should I Ask: A Knowledge-driven Approach for Follow-up Questions Generation in Conversational Surveys*
* Phillip Schneider, Nils Rehtanz, Kristiina Jokinen and Florian Matthes. *From Data to Dialogue: Leveraging the Structure of Knowledge Graphs for Conversational Exploratory Search*
* Suvodip Dey, Maunendra Sankar Desarkar, Asif Ekbal and Srijith P. K.. *DialoGen: Generalized Long-Range Context Representation for Dialogue*

3b **Syntax 1**

* Tsy Yih and Zheyuan Dai. *UPOS-DEPREL Mismatches: Detecting Annotation Errors and Improving UD Guidelines based on Linguistic Knowledge*
* Ruochen Niu, Yaqin Wang and Haitao Liu. *The Cross-linguistic Variations in Dependency Distance Minimization and its Potential Explanations*
* Zheyuan Dai, Chenliang Zhou and Haitao Liu. *Information Compression via Eliding Verb Phrase: A Dependency-Based Study*

11:45 am-1 pm Parallel Sessions 4

4a **Information Extraction**

* Dong Pham, Xanh Ho, Quang Thuy Ha and Akiko Aizawa. *Solving Label Variation in Scientific Information Extraction via Multi-Task Learning*
* Shubhangi Dutta, Manish Shrivastava and Prabhakar Bhimalapuram. *The WEAVE 2.0 Corpus: Role Labelled Synthetic Chemical Procedures from Patents with Chemical Named Entities*
* Hiranmai Adibhatla, Pavan Baswani and Manish Shrivastava. *Fine-grained Contract NER using instruction based mode*

4b **Corpus Annotation and NLP for Digital Humanities**

* Ryan Ka Yau Lai, Lily Zihe Yin, Alice Yimeng Zhang, Yuting Jiang, Bill Shiyang Xin and Junwei Gao. *Turn design, resonance and epistemic stance in the Diamond Sutra: A dialogic constructionist approach*
* Hanjie Zhao, Jinge Xie, Yuchen Yan, Yuxiang Jia, Yawen Ye and Hongying Zan. *A Corpus for Named Entity Recognition in Chinese Novels with Multi-genres*
* Ryan Ka Yau Lai. *From annotation to analysis: Exploring conversational dynamics with rezonateR*

1 pm–2:30 pm Lunch Break

2:30-3:45 pm Parallel Sessions 5

5a **Corpus Linguistics and Discourse Analysis 2**

* Sanae Yamashita, Koji Inoue, Ao Guo, Shota Mochizuki, Tatsuya Kawahara and Ryuichiro Higashinaka. *RealPersonaChat: A Realistic Persona Chat Corpus with Interlocutors' Own Personalities*
* Gregorio Ebron Jr., Romualdo Mabuan and Shirley Dita. *A Corpus-Based Investigation of Occurrences and Functions of Frame Markers in Filipino University Students’ COVID-19 Written Narratives*
* Youyou Fu. *National Identity Performance in Public Health Emergencies: A Discourse Analysis of American News Reports*

5b **NLP for Language Teaching/Learning**

* Zhendong Du and Kenji Hashimoto. *Data Augmentation for SentRev using Back-Translation of Lexical Bundles*
* Tong Zhu, Derek Irwin, Yanhui Zhang, Renjie Wu and Xiaoyi Jiang. *Estimating the Likelihood of Words Being Known with Corpus Analysis and K-Means Clustering Algorithm*
* Aleksei Abramov, Vladimir Ivanov and Valery Solovyev. *Estimating Lexical Complexity in Multi-Domain Settings for the Russian Language*

3:45 pm-4:00 pm Coffee Break

4:00 pm-5:15 pm Poster Session

* Kosuke Nishida, Naoki Yoshinaga and Kyosuke Nishida. *Sparse Neural Retrieval Model for Efficient Cross-Domain Retrieval-Based Question Answering*
* Ryutaro Asahara, Masaki Takahashi, Chiho Iwahashi and Michimasa Inaba. *SumRec: A Framework for Recommendation using Open-Domain Dialogue*
* Beatrice Portelli, Simone Scaboro and Giuseppe Serra. *Improving Multi-lingual Medical Term Normalization to Address the Long-Tail Problem*
* Masatoshi Otake and Yusuke Miyao. *Comprehensive Evaluation of Translation Error Correction Models*
* Woohyeon Moon, Taeyoung Kim, Bumgeun Park and Dongsoo Har. *Enhanced Transformer Architecture for Natural Language Processing*
* Tomoki Ikoma and Shigeki Matsubara. *Paper Recommendation Using Citation Contexts in Scholarly Documents*
* Won Ik Cho, Sangwhan Moon and Youngsook Song. *Revisiting Korean Corpus Studies through Technological Advances*
* Teruno Kajiura, Shiho Takano and Kimio Kuramitsu. *Vocabulary Replacement in SentencePiece for Domain Adaptation*
* Kazuki Murakami and Asuka Terai. *Lyrics Generation Applying Metaphor Generation*
* Tinh Pham Phuc Do, Ngoc Dinh Duy Cao, Kiet Van Nguyen, Nhan Thanh Nguyen and Tin Van Huynh. *Machine Reading Comprehension for Vietnamese Customer Reviews: Task, Corpus and Baseline Models*
* Kyosuke Takahagi and Hiroyuki Shinnou. *Data Augmentation by Shuffling Phrases in Recognizing Textual Entailment*
* Soma Asada, Kanako Komiya and Masayuki Asahara. *All-Words Word Sense Disambiguation for Historical Japanese*
* Pin-Er Chen, Po-Ya Angela Wang, Hsin-Yu Chou, Yu-Hsiang Tseng and Shu-Kai Hsieh. *Exploring Affordance and Situated Meaning in Image Captions: A Multimodal Analysis*
* Junghwan Maeng, Jinghang Gu and Sun-A Kim. *Effectiveness of ChatGPT in Korean Grammatical Error Correction*
* Toshiki Kawamoto, Yuki Okano, Takato Yamazaki, Toshinori Sato, Kotaro Funakoshi and Manabu Okumura. *A Follow-up Study on Evaluation Metrics Using Follow-up Utterances*
* Xiaotian Wang, Takuya Tamura, Takehito Utsuro and Masaaki Nagata. *Enhanced Retrieve-Edit-Rerank Framework with kNN-MT*
* Nguyen Xuan Vu Nguyen, Quang Huy Ngo and Minh Quang Nhat Pham. *Automatic Transcript Generation from Presentation Slides*
* Viktoriia Chekalina, Georgiy Novikov, Julia Gusak, Alexander Panchenko and Ivan Oseledets. Efficient GPT Model Pre-training using Tensor Train Matrix Representation
* Yuki Zenimoto, Shinzan Komata and Takehito Utsuro. *Large Scale Evaluation of End-to-End Pipeline of Speaker to Dialogue Attribution in Japanese Novels*
* Guangming Huang, Yunfei Long, Cunjin Luo and Yingya Li. *LIDA: Lexical-Based Imbalanced Data Augmentation for Content Moderation*
* Michimasa Inaba. *Generating Character Lines in Four-Panel Manga*
* Yu-Leng Lin and Jiun-Shiung Wu. *Sibilant production in Taiwan Mandarin: untangling the effects of linguistic and social variables*
* Akira Yoshida, Chihaya Matsuhira, Hirotaka Kato, Takatsugu Hirayama, Takahiro Komamizu and Ichiro Ide. *Discovering Phonesthemic Clusters in Readings of Kanji Characters toward Exploring Phonestheme in Japanese (IN PERSON)*
* Taisei Enomoto, Tosho Hirasawa, Hwichan Kim, Teruaki Oka and Mamoru Komachi. *Simultaneous Domain Adaptation of Tokenization and Machine Translation*
* Bernard Barruga. *Conceptualizing Philippine Language Policy using Liberal Neutrality Philosophical Model: A Policy Brief*
* Yishan Huang. *Encoding Syllables in Zhangzhou Southern Min – theoretical linguistics*
* Masanori Oya. *Low-Frequency Long-Distance Dependencies as “Long Tails”*
* Souhaila Khamlichi. *Redefining the Status of Moroccan Mother Tongues in Modern Morocco: The Case of Darija*
* Katsunori Kotani and Takehiko Yoshimi. *The Association of Second Language Proficiency with Nonverbal Behaviors*
* Thanh-Do Nguyen, Chi Minh Bui, Thi-Hai-Yen Vuong and Xuan-Hieu Phan. *Passage-based BM25 Hard Negatives: A Simple and Effective Negative Sampling Strategy For Dense Retrieval*
* Zhicheng Fang, Masaki Murata and Shigeki Matsubara. *Automatic Insertion of Commas and Linefeeds into Lecture Transcripts Based on Multi-task Learning*
* Aileen Bautista and Rochelle Irene Lucas. *Critical Thinking Skills in an Oral Reading Comprehension Test: A Curious Case of the Individuals with Tourette Syndrome*

5:15 pm-6:30 pm Parallel Sessions 6

6a **NLP for Dialogue and Conversation 2**

* Juntao Wang and Tsunenori Mine. *Multi-Task Learning for Emotion Recognition in Conversation with Emotion Shift*
* Yan Pan, Davide Cadamuro and Georg Groh. *Data-Augmented Task-Oriented Dialogue Response Generation with Domain Adaptation*
* Jing Yang Lee, Kong Aik Lee and Woon-Seng Gan. *Partially Randomizing Transformer Weights for Dialogue Response Diversity*

**6b Speech**

* Eran Kaufman, Lee-Ad Gottlieb, Dina Mayzlish, Or Tiram, Hila Wiesel and Nofar Yosef. *Using Deepfake Technologies for Word Emphasis Detection*
* Raviraj Joshi and Nikesh Garera. *Rapid Speaker Adaptation in Low Resource Text to Speech Systems using Synthetic Data and Transfer learning*
* 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 and Dang Huynh. *ViASR: A Novel Benchmark Dataset and Methods for Vietnamese Automatic Speech Recognition*

**Day 3: Mon, Dec 4**

9:15-10:15 am Keynote Speaker 2: Prof. Charles Yang (University of Pennsylvania)

10:15-10:30 am Coffee Break

10:30-11:45 am Parallel Sessions 7

**7a Semantics**

* Zhidong Ling, Taichi Aida, Teruaki Oka and Mamoru Komachi. *Construction of Evaluation Dataset for Japanese Lexical Semantic Change Detection*
* Yu-Hsiang Tseng, Mao-Chang Ku, Wei-Ling Chen, Yu-Lin Chang and Shu-Kai Hsieh. *Vec2Gloss: definition modeling leveraging contextualized vectors with Wordnet gloss*
* Hayate Funakura and Koji Mineshima, *Computational Semantics and Evaluation Benchmark for Interrogative Sentences via Combinatory Categorial Grammar*

**7b Sentiment and Emotion Analysis**

* Shi Yin Hong and Susan Gauch. *Improving Cross-Domain Hate Speech Generalizability with Emotion Knowledge*
* Cam-Van Nguyen Thi, Ngoc-Hoa Nguyen Thi, Duc-Trong Le and Quang-Thuy Ha. *Self-MI: Efficient Multimodal Fusion via Self-Supervised Multi-Task Learning with Auxiliary Mutual Information Maximization*
* Md Fahim, Dr. Amin Ahsan Ali, Md Ashraful Amin and Akm Mahbubur Rahman. *EDAL: Entropy based Dynamic Attention Loss for Hate Speech Classification*

11:45 am-1 pm Parallel Sessions 8

**8a Corpus Linguistics and Discourse Analysis 3**

* Grégoire Winterstein, David Vergnaud, Hannah Hoi Tung Yu, Jérémie Lupien, Laperle Samuel, Pei Sui Luk and Christopher Davis. *An empirical, corpus-based, approach to Cantonese nominal expressions*
* Naoki Nakamata. *Building a Topic-Oriented Corpus and Its Application for Language Teaching*
* Yoshiko Kawabata, Mai Omura, Masayuki Asahara and Johane Takeuchi. *Spatial Information Annotation Based on the Double Cross Model*

**8b Language Models 2**

* Kosuke Takahashi, Takahiro Omi, Kosuke Arima and Tatsuya Ishigaki. *Training Generative Question-Answering on Synthetic Data Obtained from an Instruct-tuned Model*
* Chengguang Gan and Tatsunori Mori. *Sensitivity and Robustness of Large Language Models to Prompt Template in Japanese Text Classification Tasks*
* Manh Tran-Tien, Huu-Loi Le, Dang Nhat Minh, T. Tran Khang, Huy-The Vu and Nguyen Minh-Tien. *ViPubmedDeBERTa: A Pre-trained Model for Vietnamese Biomedical Text*

1 pm–2:30 pm Lunch Break

2:30-3:45 pm Parallel Sessions 9

**9a Syntax 2**

* Jiannan Mao, Chenchen Ding, Hour Kaing, Hideki Tanaka, Masao Utiyama and Tadahiro Matsumoto. *Improving Zero-Shot Dependency Parsing by Unsupervised Learning*
* Yonghui Xie, Ruochen Niu and Haitao Liu. *Contrasting the Chinese bei-passive and the English be-passive through dependency distance*
* Huei-Ling Lin. *Directional Constructions in Taiwan Southern Min: Complex Predicates or Compounds*

**9b NLP for Low-Resource Languages 1**

* Pankaj Choudhury, Prithwijit Guha and Sukumar Nandi. *Image Caption Synthesis for Low Resource Assamese Language using Bi-LSTM with Bilinear Attention*
* Ma. Beatrice Emanuela Pilar, Dane Dedoroy, Ellyza Mari Papas, Mary Loise Buenaventura, Myron Darrel Montefalcon, Jay Rhald Padilla, Joseph Marvin Imperial, Mideth Abisado and Lany Maceda. *CebuaNER: A New Baseline Cebuano Named Entity Recognition Model*
* Aabha Pingle, Aditya Vyawahare, Isha Joshi, Rahul Tangsali and Raviraj Joshi. *L3Cube-MahaSent-MD: A Multi-domain Marathi Sentiment Analysis Dataset and Transformer Models*

3:45 pm-4:00 pm Coffee Break

4:00 pm-5:15 pm Parallel Sessions 10

**10a NLP for Low Resource Languages 2**

* Kasun Wickramasinghe and Nisansa de Silva. *Sinhala-English Word Embedding Alignment: Introducing Datasets and Benchmark for a Low Resource Language*
* Kyle Chan, Kaye Ann De Las Alas, Charles Orcena, Dan John Velasco, Qyle John San Juan and Charibeth Cheng. *Practical Approaches for Low-Resource Named Entity Recognition of Filipino Telecommunications Domain*
* Hemanta Baruah, Sanasam Ranbir Singh and Priyankoo Sarmah. *Assamese Back Transliteration - An Empirical Study Over Canonical and Non-canonical Datasets*

**10b Multilingual NLP**

* Telem Joyson Singh, Sanasam Ranbir Singh and Priyankoo Sarmah. *Subwords to Word Back Composition for Morphologically Rich Languages in Neural Machine Translation*
* Samruddhi Deode, Janhavi Gadre, Aditi Kajale, Ananya Joshi and Raviraj Joshi. *L3Cube-IndicSBERT: A simple approach for learning cross-lingual sentence representations using multilingual BERT*
* Lokesh Madasu, Gopichand Kanumolu, Nirmal Surange and Manish Shrivastava. *Mukhyansh: A Headline Generation Dataset for Indic Languages*

5:15-5:45 pm Invited Talk: Furhat Team

5:45 pm-6:00 pm Conclusive Remarks

7:30 Social Dinner at the Nan Lian Garden Vegetarian Restaurant