



2:15 - 2:30 pm
2:30 - 3:15 pm

Introduction

Invited Talk 1 (TU201)

Dr. Enrico Santus (Bloomberg) Chair: Chu-Ren Huang (The Hong Kong Polytechnic University, Hong Kong)
Human-AI Collaboration & Adaptive Processes in Industry

Meeting ID: 587 510 3983

Zoom link: <https://polyu.zoom.us/j/5875103983>

3:15 - 4:30 pm

1a

Parallel Sessions 1

(TU 103) Psycholinguistics and Experimental Linguistics (Chair: James Britton, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 901 141 0734

Zoom link: <https://polyu.zoom.us/j/9011410734>

Japanese benefactive passives are more difficult to comprehend than benefactive actives

- Masataka Ogawa.

Word Familiarity Rate Estimation for Japanese Functional Words Using a Bayesian Linear Mixed Model

- Bocheng Chen and Masayuki Asahara.

A Methodology for Exploring Experiences of Individuals with Tourette Syndrome in Oral and Silent Reading Assessments

- Aileen Bautista, Macario li Cordel, Cely Magpantay and Catherine Manuela Ramos.

1b

(TU 201) Language Models 1 (Chair: Jinghang Gu, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 287 105 8299

Zoom link: <https://polyu.zoom.us/j/2871058299>

Large Language Models Meet Knowledge Graphs to Answer Factoid Questions

- Mikhail Salnikov, Hai Le, Prateek Rajput, Irina Nikishina, Pavel Braslavski, Valentin Malykh and Alexander Panchenko.

Counterfactual Adversarial Training for Improving Robustness of Pre-trained Language Models

- Hoai Linh Luu and Naoya Inoue.

A computational study of matrix decomposition methods for compression of pre-trained Transformers

- Sergey Pletenev, Viktoriia Chekalina, Daniil Moskovskiy, Mikhail Seleznev, Sergey Zagoruyko and Alexander Panchenko.

4:30 - 4:45 pm

Coffee Break

4:45 - 6:00 pm

2a

Parallel Sessions 2

(TU 103) Corpus Linguistics and Discourse Analysis 1 (Chair: Gabor Parti, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 901 141 0734

Zoom link: <https://polyu.zoom.us/j/9011410734>

Toward a corpus-based identification of nominal relationality and uniqueness: A constructionist approach

- Kazuho Kambara and Taishi Chika.

Tracing Social Change through Metaphor: A Diachronic Corpus-Assisted Analysis

- Winnie Huiheng Zeng, Kathleen Ahrens and Chu-Ren Huang.

Constructing a Japanese Business Email Corpus Based on Social Situations

- Muxuan Liu, Tatsuya Ishigaki, Yusuke Miyao, Hiroya Takamura and Ichiro Kobayashi.

2b

(TU201) NLP for Social Media (Chair: He Zhou, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 287 105 8299

Zoom link: <https://polyu.zoom.us/j/2871058299>

Automatic Identification of Assistance Needs in Disaster Situations Using Hybrid Word Embedding Techniques

- Sarra Chairir, Malika Charrad and Narjès Bellamine Ben Saoud.

Developing an Annotated Persian Dataset from COVID-19 News for Enhanced Fake News Detection

- Foroogh Zahed, Seyedeh Fatemeh Ebrahimi, Mohammad Bahrani and Alireza Mansouri.

Linguistic and Paralinguistic Features influencing Reliability Judgments of Thai Twitter Reviews

- Warisaraporn Limprasert, Nanthicha Angsuwichitkul and Francesco Burrone.

Data Augmentation for SentRev using Back-Translation of Lexical Bundles

- Zhendong Du and Kenji Hashimoto. Data Augmentation for SentRev using Back-Translation of Lexical Bundles

7:00 pm

Welcome Reception at Dim Sum Restaurant



9:15 - 10:15 am

Keynote Speaker 1: (TU 201)

Prof. Lori Levin (Carnegie Mellon University) (Chair: Kathleen Ahrens, The Hong Kong Polytechnic University, Hong Kong)

Interlinear Gloss in Natural Language Processing and Language Documentation

Meeting ID: 587 510 3983

Zoom link: <https://polyu.zoom.us/j/5875103983>

10:15 - 10:30 am

Coffee Break

10:30 - 11:45 am

Parallel Sessions 3

3a

(TU 107) NLP for Dialogue and Conversation 1 (Chair: Giuseppe Serra, University of Udine, Italy)

Meeting ID: 901 141 0734

Zoom link: <https://polyu.zoom.us/j/9011410734>

What should I Ask: A Knowledge-driven Approach for Follow-up Questions Generation in Conversational Surveys

- Yubin Ge, Ziang Xiao, Jana Diesner, Heng Ji, Karrie Karahalios and Hari Sundaram.

From Data to Dialogue: Leveraging the Structure of Knowledge Graphs for Conversational Exploratory Search

- Phillip Schneider, Nils Rehtanz, Kristiina Jokinen and Florian Matthes.

DialoGen: Generalized Long-Range Context Representation for Dialogue

- Suvodip Dey, Maunendra Sankar Desarkar, Asif Ekbal and Srijith P. K..

3b

(TU 201) Syntax 1 (Chair: Henrison Hsieh, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 287 105 8299

Zoom link: <https://polyu.zoom.us/j/2871058299>

UPOS-DEPREL Mismatches: Detecting Annotation Errors and Improving UD Guidelines based on Linguistic Knowledge

- Tsy Yih and Zheyuan Dai.

The Cross-linguistic Variations in Dependency Distance Minimization and its Potential Explanations

- Ruochen Niu, Yaqin Wang and Haitao Liu.

Information Compression via Eliding Verb Phrase: A Dependency-Based Study

- Zheyuan Dai, Chenliang Zhou and Haitao Liu.

11:45 am - 1 pm

4a

Parallel Sessions 4

(TU 107) Information Extraction (Chair: Jinghang Gu, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 901 141 0734

Zoom link: <https://polyu.zoom.us/j/9011410734>

Solving Label Variation in Scientific Information Extraction via Multi-Task Learning

- Dong Pham, Xanh Ho, Quang Thuy Ha and Akiko Aizawa.

The WEAVE 2.0 Corpus: Role Labelled Synthetic Chemical Procedures from Patents with Chemical Named Entities

- Shubhangi Dutta, Manish Shrivastava and Prabhakar Bhimalapuram.

Fine-grained Contract NER using instruction based mode

- Hiranmai Adibhatla, Pavan Baswani and Manish Shrivastava.

t

4b

(TU 201) Corpus Annotation and NLP for Digital Humanities (Chair: Sophia Lee, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 287 105 8299

Zoom link: <https://polyu.zoom.us/j/2871058299>

Turn design, resonance and epistemic stance in the Diamond Sutra: A dialogic constructionist approach

- Ryan Ka Yau Lai, Lily Zihe Yin, Alice Yimeng Zhang, Yuting Jiang, Bill Shiyang Xin and Junwei Gao.

A Corpus for Named Entity Recognition in Chinese Novels with Multi-genres

- Hanjie Zhao, Jingjie Xie, Yuchen Yan, Yuxiang Jia, Yawen Ye and Hongying Zan.

From annotation to analysis: Exploring conversational dynamics with resonateR

- Ryan Ka Yau Lai.

1:00 - 2:30 pm

Lunch Break

2:30 - 3:45 pm
5a**Parallel Sessions 5****(TU 107) Corpus Linguistics and Discourse Analysis 2** (Chair: Winnie Huiheng Zeng, Hong Kong Shue Yan University, Hong Kong)**Meeting ID: 901 141 0734****Zoom link: <https://polyu.zoom.us/j/9011410734>****RealPersonaChat: A Realistic Persona Chat Corpus with Interlocutors' Own Personalities**

- Sanae Yamashita, Koji Inoue, Ao Guo, Shota Mochizuki, Tatsuya Kawahara and Ryuichiro Higashinaka.

A Corpus-Based Investigation of Occurrences and Functions of Frame Markers in Filipino University Students' COVID-19 Written Narratives

- Gregorio Ebron Jr., Romualdo Mabuan and Shirley Dita.

National Identity Performance in Public Health Emergencies: A Discourse Analysis of American News Reports

- Youyou Fu.

5b

(TU 201) NLP for Language Teaching/Learning (Chair: Won-Ik Cho, Samsung Advanced Institute of Technology, South Korea)**Meeting ID: 287 105 8299****Zoom link: <https://polyu.zoom.us/j/2871058299>****Effectiveness of ChatGPT in Korean Grammatical Error Correction**

- Junghwan Maeng, Jinghang Gu and Sun-A Kim.

Estimating the Likelihood of Words Being Known with Corpus Analysis and K-Means Clustering Algorithm

- Tong Zhu, Derek Irwin, Yanhui Zhang, Renjie Wu and Xiaoyi Jiang.

Estimating Lexical Complexity in Multi-Domain Settings for the Russian Language

- Aleksei Abramov, Vladimir Ivanov and Valery Solovyev.

3:45 - 4:00 pm

Coffee Break

4:00 - 5:15 pm

Poster Session**Sparse Neural Retrieval Model for Efficient Cross-Domain Retrieval-Based Question Answering**

- Kosuke Nishida, Naoki Yoshinaga and Kyosuke Nishida.

SumRec: A Framework for Recommendation using Open-Domain Dialogue

- Ryutarō Asahara, Masaki Takahashi, Chiho Iwahashi and Michimasa Inaba.

Improving Multi-lingual Medical Term Normalization to Address the Long-Tail Problem

- Beatrice Portelli, Simone Scaboro and Giuseppe Serra.

Comprehensive Evaluation of Translation Error Correction Models

- Masatoshi Otake and Yusuke Miyao.

Paper Recommendation Using Citation Contexts in Scholarly Documents

- Tomoki Ikoma and Shigeki Matsubara.

Revisiting Korean Corpus Studies through Technological Advances

- Won Ik Cho, Sangwhan Moon and Youngsook Song.

Vocabulary Replacement in SentencePiece for Domain Adaptation

- Teruno Kajiura, Shiho Takano, Tatsuya Hiraoka, and Kimio Kuramitsu.

Data Augmentation by Shuffling Phrases in Recognizing Textual Entailment

- Kyosuke Takahagi and Hiroyuki Shinnou.

All-Words Word Sense Disambiguation for Historical Japanese

- Soma Asada, Kanako Komiya and Masayuki Asahara.

Exploring Affordance and Situated Meaning in Image Captions: A Multimodal Analysis

- Pin-Er Chen, Po-Ya Angela Wang, Hsin-Yu Chou, Yu-Hsiang Tseng and Shu-Kai Hsieh.



4:00 - 5:15 pm

Poster Session (Continued)**A Follow-up Study on Evaluation Metrics Using Follow-up Utterances**

- Toshiki Kawamoto, Yuki Okano, Takato Yamazaki, Toshinori Sato, Kotaro Funakoshi and Manabu Okumura.

Enhanced Retrieve-Edit-Rerank Framework with kNN-MT

- Xiaotian Wang, Takuya Tamura, Takehito Utsuro and Masaaki Nagata.

Automatic Transcript Generation from Presentation Slides

- Nguyen Xuan Vu Nguyen, Quang Huy Ngo and Minh Quang Nhat Pham.

Efficient GPT Model Pre-training using Tensor Train Matrix Representation

- Viktoriia Chekalina, Georgiy Novikov, Julia Gusak, Alexander Panchenko and Ivan Oseledets.

Large Scale Evaluation of End-to-End Pipeline of Speaker to Dialogue Attribution in Japanese Novels

- Yuki Zenimoto, Shinzan Komata and Takehito Utsuro.

LIDA: Lexical-Based Imbalanced Data Augmentation for Content Moderation

- Guangming Huang, Yunfei Long, Cunjin Luo and Yingya Li.

Generating Character Lines in Four-Panel Manga

- Michimasa Inaba.

Sibilant production in Taiwan Mandarin: untangling the effects of linguistic and social variables

- Yu-Leng Lin and Jiun-Shiung Wu.

Discovering Phonesthemic Clusters in Readings of Kanji Characters toward Exploring Phonestheme in Japanese

- Akira Yoshida, Chihaya Matsuhira, Hirota Kato, Takatsugu Hirayama, Takahiro Komamizu and Ichiro Ide.

Simultaneous Domain Adaptation of Tokenization and Machine Translation

- Taisei Enomoto, Toshio Hirasawa, Hwihan Kim, Teruaki Oka and Mamoru Komachi.

The Association of Second Language Proficiency with Nonverbal Behaviors

- Katsunori Kotani and Takehiko Yoshimi.

Automatic Insertion of Commas and Linefeeds into Lecture Transcripts Based on Multi-task Learning

- Zhicheng Fang, Masaki Murata and Shigeki Matsubara.

Critical Thinking Skills in an Oral Reading Comprehension Test: A Curious Case of the Individuals with Tourette Syndrome

- Aileen Bautista and Rochelle Irene Lucas.

4:00 - 5:15 pm

VIRTUAL POSTER SESSION: NLP (Chairs: Junlin Li and Yu xi Li, The Hong Kong Polytechnic University, Hong Kong)**Meeting ID: 901 141 0734**Zoom link: <https://polyu.zoom.us/j/9011410734>**Machine Reading Comprehension for Vietnamese Customer Reviews: Task, Corpus and Baseline Models**

- Tinh Pham Phuc Do, Ngoc Dinh Duy Cao, Kiet Van Nguyen, Nhan Thanh Nguyen and Tin Van Huynh.

Enhanced Transformer Architecture for Natural Language Processing

- Woohyeon Moon, Taeyoung Kim, Bumgeun Park and Dongsoo Har.

Lyrics Generation Applying Metaphor Generation

- Kazuki Murakami and Asuka Terai.

Passage-based BM25 Hard Negatives: A Simple and Effective Negative Sampling Strategy For Dense Retrieval

- Thanh-Do Nguyen, Chi Minh Bui, Thi-Hai-Yen Vuong and Xuan-Hieu Phan.



4:00 - 5:15 pm

VIRTUAL POSTER SESSION: LINGUISTICS (Chairs: Jing Chen and Gabor Parti, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 287 105 8299

Zoom link: <https://polyu.zoom.us/j/2871058299>

Conceptualizing Philippine Language Policy using Liberal Neutrality Philosophical Model: A Policy Brief

- Bernard Barruga.

Encoding Syllables in Zhangzhou Southern Min

- Yishan Huang.

Low-Frequency Long-Distance Dependencies as “Long Tails”

- Masanori Oya.

Redefining the Status of Moroccan Mother Tongues in Modern Morocco: The Case of Darija

- Souhaila Khamlichi.

5:15 - 6:30 pm

Parallel Sessions 6

(TU107) NLP for Dialogue and Conversation 2 (Chair: Bo Peng, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 901 141 0734

Zoom link: <https://polyu.zoom.us/j/9011410734>

(TU 107) Multi-Task Learning for Emotion Recognition in Conversation with Emotion Shift

- Juntao Wang and Tsunenori Mine.

Data-Augmented Task-Oriented Dialogue Response Generation with Domain Adaptation

- Yan Pan, Davide Cadamuro and Georg Groh.

Partially Randomizing Transformer Weights for Dialogue Response Diversity

- Jing Yang Lee, Kong Aik Lee and Woon-Seng Gan.

6b

(TU 201) Speech (Chair: Si Chen, The Hong Kong Polytechnic University, Hong Kong)

Meeting ID: 287 105 8299

Zoom link: <https://polyu.zoom.us/j/2871058299>

Using Deepfake Technologies for Word Emphasis Detection

- Eran Kaufman, Lee-Ad Gottlieb, Dina Mayzlish, Or Tiram, Hila Wiesel and Nofar Yosef.

Rapid Speaker Adaptation in Low Resource Text to Speech Systems using Synthetic Data and Transfer learning

- Raviraj Joshi and Nikesh Garera.

ViASR: A Novel Benchmark Dataset and Methods for Vietnamese Automatic Speech Recognition

- 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.

9:15 - 10:15 am

Keynote Speaker 2: (TU 201)**Prof. Charles Yang (University of Pennsylvania)** (Chair: Emmanuele Chersoni, The Hong Kong Polytechnic University, Hong Kong)**Learning by Satisficing****Meeting ID: 587 510 3983**Zoom link: <https://polyu.zoom.us/j/5875103983>

10:15 - 10:30 am

Coffee Break

10:30 - 11:45 am

Parallel Sessions 7

7a

(TU 201) Semantics (Chair: Emmanuele Chersoni, The Hong Kong Polytechnic University, Hong Kong)**Meeting ID: 901 141 0734**Zoom link: <https://polyu.zoom.us/j/9011410734>**Construction of Evaluation Dataset for Japanese Lexical Semantic Change Detection**

- Zhidong Ling, Taichi Aida, Teruaki Oka and Mamoru Komachi.

Vec2Gloss: definition modeling leveraging contextualized vectors with Wordnet gloss

- Yu-Hsiang Tseng, Mao-Chang Ku, Wei-Ling Chen, Yu-Lin Chang and Shu-Kai Hsieh.

Computational Semantics and Evaluation Benchmark for Interrogative Sentences via Combinatory Categorical Grammar

- Hayate Funakura and Koji Mineshima.

7b

(TU 107) Sentiment and Emotion Analysis (Chair: Lavinia Salicchi, The Hong Kong Polytechnic University, Hong Kong)**Meeting ID: 287 105 8299**Zoom link: <https://polyu.zoom.us/j/2871058299>**Improving Cross-Domain Hate Speech Generalizability with Emotion Knowledge**

- Shi Yin Hong and Susan Gauch.

Self-MI: Efficient Multimodal Fusion via Self-Supervised Multi-Task Learning with Auxiliary Mutual Information Maximization

- Cam-Van Nguyen Thi, Ngoc-Hoa Nguyen Thi, Duc-Trong Le and Quang-Thuy Ha.

EDAL: Entropy based Dynamic Attention Loss for Hate Speech Classification

- Md Fahim, Dr. Amin Ahsan Ali, Md Ashraful Amin and Akm Mahbubur Rahman.

11:45 am - 1 pm

8a

Parallel Sessions 8**(TU 201) Corpus Linguistics and Discourse Analysis 3** (Chair: Mohammad Momenian, The Hong Kong Polytechnic University, Hong Kong)**Meeting ID: 901 141 0734**Zoom link: <https://polyu.zoom.us/j/9011410734>**An empirical, corpus-based, approach to Cantonese nominal expressions**

- Grégoire Winterstein, David Vergnaud, Hannah Hoi Tung Yu, Jérémie Lupien, Laperle Samuel, Pei Sui Luk and Christopher Davis.

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

- Naoki Nakamata.

Spatial Information Annotation Based on the Double Cross Model

- Yoshiko Kawabata, Mai Omura, Masayuki Asahara and Johane Takeuchi.

8b

(TU 107) Language Models 2 (Chair: Bo Peng, The Hong Kong Polytechnic University, Hong Kong)**Meeting ID: 287 105 8299**Zoom link: <https://polyu.zoom.us/j/2871058299>**Training Generative Question-Answering on Synthetic Data Obtained from an Instruct-tuned Model**

- Kosuke Takahashi, Takahiro Omi, Kosuke Arima and Tatsuya Ishigaki.

Sensitivity and Robustness of Large Language Models to Prompt Template in Japanese Text Classification Tasks

- Chengguang Gan and Tatsunori Mori.

ViPubMedDeBERTa: A Pre-trained Model for Vietnamese Biomedical Text

- Manh Tran-Tien, Huu-Loi Le, Dang Nhat Minh, T. Tran Khang, Huy-The Vu and Nguyen Minh-Tien.

1:00 - 2:30 pm

Lunch Break

2:30 - 3:45 pm
9a**Parallel Sessions 9****(TU 201) Syntax 2** (Chair: Henrison Hsieh, The Hong Kong Polytechnic University, Hong Kong)**Meeting ID: 901 141 0734**Zoom link: <https://polyu.zoom.us/j/9011410734>**Improving Zero-Shot Dependency Parsing by Unsupervised Learning**

- Jiannan Mao, Chenchen Ding, Hour Kaing, Hideki Tanaka, Masao Utiyama and Tadahiro Matsumoto.

10:15 - 10:30 am

Contrasting the Chinese bei-passive and the English be-passive through dependency distance

- Yonghui Xie, Ruochen Niu and Haitao Liu.

10:30 - 11:45 am

Directional Constructions in Taiwan Southern Min: Complex Predicates or Compounds

- Huei-Ling Lin.

9b

(TU 107) NLP for Low-Resource Languages 1 (Chair: Pranav A, Dayta AI, Hong Kong)**Meeting ID: 287 105 8299**Zoom link: <https://polyu.zoom.us/j/2871058299>**Image Caption Synthesis for Low Resource Assamese Language using Bi-LSTM with Bilinear Attention**

- Pankaj Choudhury, Prithwjit Guha and Sukumar Nandi.

CebuNER: A New Baseline Cebuano Named Entity Recognition Model

- Ma. Beatrice Emanuela Pilar, Dane Dedoroy, Ellyza Mari Papas, Mary Loise Buenaventura, Myron Darrel Montefalcon, Jay Rhalad Padilla, Joseph Marvin Imperial, Mideth Abisado and Lany Maceda.

L3Cube-MahaSent-MD: A Multi-domain Marathi Sentiment Analysis Dataset and Transformer Models

- Aabha Pingle, Aditya Vyawahare, Isha Joshi, Rahul Tangsali and Raviraj Joshi.

3:45 - 4:00 pm

Coffee Break

4:00 - 5:15 pm

10a

Parallel Sessions 10**(TU 201) NLP for Low Resource Languages 2** (Chair: Jing Li, The Hong Kong Polytechnic University, Hong Kong)**Meeting ID: 901 141 0734**Zoom link: <https://polyu.zoom.us/j/9011410734>**Sinhala-English Word Embedding Alignment: Introducing Datasets and Benchmark for a Low Resource Language**

- Kasun Wickramasinghe and Nisansa de Silva.

Practical Approaches for Low-Resource Named Entity Recognition of Filipino Telecommunications Domain

- Kyle Chan, Kaye Ann De Las Alas, Charles Orcena, Dan John Velasco, Qyle John San Juan and Charibeth Cheng.

Assamese Back Transliteration - An Empirical Study Over Canonical and Non-canonical Datasets

- Hemanta Baruah, Sanasam Ranbir Singh and Priyankoo Sarmah.

10b

(TU 107) Multilingual NLP (Chair: Pranav A, Dayta AI, Hong Kong)**Meeting ID: 287 105 8299**Zoom link: <https://polyu.zoom.us/j/2871058299>**Subwords to Word Back Composition for Morphologically Rich Languages in Neural Machine Translation**

- Telem Joyson Singh, Sanasam Ranbir Singh and Priyankoo Sarmah.

L3Cube-IndicSBERT: A simple approach for learning cross-lingual sentence representations using multilingual BERT

- Samruddhi Deode, Janhavi Gadre, Aditi Kajale, Ananya Joshi and Raviraj Joshi.

Mukhyansh: A Headline Generation Dataset for Indic Languages

- Lokesh Madasu, Gopichand Kanumolu, Nirmal Surange and Manish Shrivastava.

5:15 - 5:45 pm

Invited Talk 2: (TU201)**Furhat Robotics** (Chair: Bruce Wang, The Hong Kong Polytechnic University, Hong Kong)**From Breakdown to Breakthrough: Overcoming 'Robot Death' in Social Interactions****Meeting ID: 587 510 3983**Zoom link: <https://polyu.zoom.us/j/5875103983>

5:45 - 6:00 pm

Conclusive Remarks

7:30 pm

Social Dinner at the Nan Lian Garden Vegetarian Restaurant

*Octopus card is suggested for transportation.