

PACLIC35 (2021)

**Proceedings of the 35th Pacific Asia Conference on Language,
Information and Computation**

5–7 November, 2021

Shanghai International Studies University

Shanghai, China

©The PACLIC 35 Organizing Committee and PACLIC Steering Committee

All rights reserved. Except as otherwise expressly permitted under copyright law, no part of this publication may be reproduced, digitized, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, Internet or otherwise, without the prior permission of the publisher.

Copyright of contributed papers reserved by respective authors.

ISSN 2619-7782

Acknowledgement

PACLIC 35 is hosted by Shanghai International Studies University in conjunction with China University Alliance for the Development of Foreign Language and Literature Discipline.

Forword

The 35th Pacific Asia Conference on Language, Information and Computation (PACLIC 35) is organized by the Institute of Corpus Studies and Applications of Shanghai International Studies University, November 5-7, 2021. This edition of the PACLIC series of conferences, as its long tradition, continues to emphasize the synergy of theoretical analysis and processing of natural language, aiming to strengthen the communication between researchers working in different fields of language study in the Asia-Pacific region as well as around the world.

As the COVID-19 pandemic has seen dim possibility of ending in a foreseeable future globally, like the 34th conference in organization, this year's event is also organized totally online. We received 110 submissions, out of which 52 were accepted for oral presentations and 32 for poster presentations. Two accepted oral papers were withdrawn after the closing of the main conference due to expiration of the deadline for their being updated. The acceptance rate for oral presentations and poster presentations are 47% and 29% respectively. In addition to oral and poster presentations, the conference highlights five keynote talks. We are grateful to Chungmin Lee, Yue Zhang, Francesca Strik Lievers, Jingxia Lin, and Mitch Marcus for accepting to give a keynote talk. Besides, our gratitude also goes to Lily I-Wen Su, though we felt sorry to know that she had to decline an invited talk due to health problems.

There are more people to whom we'd like to express our sincere gratitude, because without their support and assistance especially in a world full of too many high contingencies, PACLIC 35 would not be made possible. Program committee members and sub-reviewers offer their professional reviews so that we can maintain the high-quality standard of PACLIC. A special thank goes to Kai Zhang from the Institute of Corpus Studies and Applications for maintaining the conference website. We are deeply indebted to the local organizing committee Kaibao Hu, Qiang Geng, Emmanuele Chersoni, as well as student staff members from the Institute of Corpus Studies and Applications of Shanghai International Studies University. Finally we would like to thank Shanghai International Studies University for sponsoring the conference so it can work properly without receiving any registration fees from all our attendees. SISU's scientific support is the guarantee indispensable for the success of this year's event.

Chu-Ren Huang
Jong-Bok Kim
Ryo Ootoguro
Rachel Edita O. Roxas
Maosong Sun
Benjamin T'sou
Min Zhang

PACLIC Steering Committee

Organizers

PACLIC Steering Committee Standing Members:

Chu-Ren Huang, The Hong Kong Polytechnic University, Hong Kong

Jong-Bok Kim, Kyung Hee University, Seoul

Ryo Ootoguro, Waseda University, Tokyo

Rachel Edita O. Roxas, National University, Manila

Maosong Sun, Tsinghua University, Beijing

Benjamin T'sou, City University of Hong Kong, Hong Kong

Min Zhang, Soochow University, Suzhou

Organizing Committee

Organizing Chair:

Kaibao Hu, Shanghai International Studies University

Co-Chairs:

Hai Zhao, Shanghai Jiao Tong University

NGUYEN Thi Minh Huyen, VNU University of Science

Program Committee

Co-Chairs:

Jong-Bok Kim, Kyung Hee University

Chengqing Zong, Institute of Automation, Chinese Academy of Science

Emmanuele Chersoni, The Hong Kong Polytechnic University

Local Members:

Boxing Chen, The Academy for Discovery, Adventure, Momentum and Outlook, Alibaba DAMO Academy

Hongwei Ding, Shanghai Jiao Tong University

Wenwei Mao, Shanghai International Studies University

Jianjun Shi, Shanghai International Studies University

Jidong Zhang, Shanghai International Studies University

Yong Jiang, Shanghai International Studies University

Huaqing Hong, Shanghai International Studies University

Local Organizing Committee

Co-Chairs:

Ziman Han, Shanghai International Studies University

Qiang Geng, Shanghai International Studies University

Members:

Hongzhi Xu, Shanghai International Studies University

Juhua Jin, Shanghai International Studies University

Xiaoqian Li, Shanghai International Studies University
Shanshan Lv, Shanghai International Studies University

Reviewers:

Shu-Ing Shyu	Pornsiri Singhapreecha	Philippe Blache
Emmanuele Chersoni	Helena Gao	Huaqing Hong
Muhammad Afzaal	Yunfei Long	Phuong Le-Hong
Te-Hsin Liu	Mai Chi Luong	Zhongyang Li
Thi Thu Trang Nguyen	Jianjun Shi	Mieradilijiang Maimaiti
Nattama Pongpairoj	Sanghoun Song	Thi Minh Huyen Nguyen
Jiun-Shiung Wu	Zhiyang Teng	Chutamane Onsuwan
Hongzhi Xu	Zhongqing Wang	Jin-Woo Chung
Yanhui Gu	Jidong Zhang	Valia Kordoni
Munpyo Hong	Hongwei Ding	Jinghang Gu
Kyunghye Kim	Xiaoqian Li	Rong Xiang
Huei-Ling Lai	Jian-Yun Nie	Tony Nguyen
Yong-Hun Lee	Tak-Sum Wong	Nguyen Thi Mung
Wenwei Mao	Cheng-Zen Yang	Dat Vi
Nathaniel Oco	Wenpeng Yin	Pham Viet Thanh
Ethel Ong	Satoru Yokoyama	Bo Peng
Jong C. Park	Liang-Chih Yu	Dang Trung Duc Anh
Melanie Siegel	Nguyen Bach	

Invited Speakers:

Chungmin Lee, Seoul National University
Zhang Yue, Westlake University
Francesca Strik Lievers, Department of Modern Languages and Literatures, University of Genoa
Lin Jingxia, School of Humanities, Nanyang Technological University Singapore
Mitch Marcus, Department of Computer and Information Science, University of Pennsylvania

Invited Talks

Chungmin Lee: Typology Of Factivity Alternation In Different Languages

This talk aims to explore the typology of factivity alternation in different languages to seek the ultimate goal of semantic universals of possible factive and non-factive alternants of epistemically-oriented attitude predicates. Korean and other Altaic languages show factivity alternation of the epistemically-oriented cognitive attitude verb ‘know’ and such verbs as ‘remember,’ ‘understand’ and ‘recognize’ between factive vs. non-factive (Lee 1978, 2019).

In contrast, Chinese rigidly reveals a typical type of non-alternating language; its cognitive epistemic verb *zhidao* ‘know’ is constantly factive and has no non-factive alternant, along with related attitude verbs such as *mingbai/lijie* ‘understand,’ *yishidao* ‘recognize,’ etc. However, the attitude verb *jide* ‘remember’ alone reveals factivity alternation with factive vs. non-factive alternant readings, with no syntactic differences in its embedded complement clauses.

English is close to a non-alternating language; the verb *know* belongs to a cognitive factive verb, although it shows exceptional non-factive readings in presupposition-cancelling contexts such as some negation, interrogative, or before.

Altaic gives some clue to alternation with its complementation typing: the thematic argument clausal DP (ACC) is embedded by the factive alternant and the REPORTative C(omp) [with covert SAY] finite mood clause is embedded by the non-factive alternant of an epistemically-oriented attitude verbs ‘know’ and ‘remember’. Such principled decomposition and compositionality may lead to semantic universals of possible structural and contextual factive and non-factive alternants of attitude predicates. Factivity is not a myth.

Yue Zhang: Guiding Document Encoding For Sequence-To-Sequence Tasks—A Soft Constraint Method

Encoding long documents can be challenging in sequence-to-sequence tasks due to much increased input length as compared with sentence level tasks. On the one hand, separately encoding each sentence has the limitation of losing useful discourse and coreference information; on the other hand, encoding the document as a single unit can lead to diluted focus in the decoder. This issue has been exemplified in neural machine translation, where it has been shown difficult to achieve training convergence using a standard Transformer for document encoding. We address the above issue by using soft constraints to guide the decoder. In particular, a set of group tags, resembling positional embedding vectors, are used to guide attention from the decoder to the encoder. For neural machine translation, we use such soft constraints to guide the current sentence being translated; for abstractive rewriting of extractive text summarization, we use constraints to guide the decoder towards the extracted content being rewritten. Results show that the soft constraints are quite effective, facilitating model convergence and leading to the best results.

Francesca Strik Lievers: Sensory language and (cross)sensory metaphors

Language allows us to talk about what we perceive, but sensory words available in the lexicon of individual languages tend to be distributed unevenly across the senses. In English, for instance, there are many words for sounds, but only few for smells. Interestingly, when we use words from one sense to describe perceptions in another sense, typically through synaesthetic metaphors, we also find asymmetries between the senses. Corpus-based studies show that in many languages it is for example common

to use a touch adjective to modify a hearing noun, as in “warm voice”, while finding a hearing adjective modifying a touch noun is less likely. In this talk, I will discuss such asymmetries in the sensory lexicon and in synaesthetic metaphors and examine possible motivations for both.

Jingxia Lin: Scale in Mandarin Chinese

This talk introduces the notion of scale and shows how it may help to better explain the syntactic and semantic features of verbs, adjectives, and preposition phrases in Modern Mandarin Chinese. The topics to be covered in this talk include the aspectual classification of verbs, the quantitative denotations of simple adjectives, and the word order of spatial preposition phrases.

Mitch Marcus: Language learning: Artificial Neural Nets vs. Real Neurons

Unsupervised learning of natural languages by neural nets requires 10^8 or more words or characters of text. But children learn within a year or two with on the order of 10^3 input tokens. How is that? ... This talk will survey a research program of my research group over many years investigating what clues children might use to learn language efficiently and how we might exploit these clues to build efficient unsupervised language learners.

Table of Contents

Regular Papers

Under the Microscope: Interpreting Readability Assessment Models for Filipino	1
<i>Joseph Marvin Imperial and Ethel Ong</i>	
SEPP: Similarity Estimation of Predicted Probabilities for Defending and Detecting Adversarial Text	11
<i>Hoang-Quoc Nguyen-Son, Seira Hidano, Kazuhide Fukushima and Shinsaku Kiyomoto</i>	
A Comprehensive Analysis of PMI-based Models for Measuring Semantic Differences	21
<i>Taichi Aida, Mamoru Komachi, Toshinobu Ogiso, Hiroya Takamura and Daichi Mochihashi</i>	
Exploring sentiment constructions: connecting deep learning models with linguistic construction .	32
<i>Shu-Kai Hsieh and Yu-Hsiang Tseng</i>	
A Neuro-Symbolic Approach for Question Answering on Research Articles	40
<i>Komal Gupta, Tirthankar Ghosal and Asif Ekbal</i>	
An Empirical Performance Analysis of State-of-the-Art Summarization Models for Automatic Minuting	50
<i>Muskaan Singh, Tirthankar Ghosal Ghosal and Ondrej Bojar</i>	
Data Augmentation to Address Out of Vocabulary Problem in Low Resource Sinhala English Neural Machine Translation	61
<i>Aloka Fernando and Surangika Ranathunga</i>	
“Prickly Voice” or “Smelly Voice”? Comprehending novel synaesthetic metaphors	71
<i>Yin Zhong and Kathleen Ahrens</i>	
Translation for professionals: corpus-based study of translation universals in computing	80
<i>Huang Baorong</i>	
Island effects in Vietnamese relativization: A formal judgment study	89
<i>Quyên Nguyen</i>	
Lexical Meanings of Ditransitive Verbs in Japanese	99
<i>Hiroko Kimura and Chigusa Morita</i>	
Examine persuasion strategies in Chinese on social media	108
<i>Yu-Yun Chang, Po-Ya Angela Wang, Han-Tang Hung, Ka-Sîng Khóo and Shu-Kai Hsieh</i>	
Exploring Conditional Text Generation for Aspect-Based Sentiment Analysis	119
<i>Siva Uday Sampreeth Chebolu, Franck Deroncourt, Nedim Lipka and Thamar Solorio</i>	
Aspect or Manner? A Study of Reduplicated Adverbials in Mandarin Chinese	130
<i>Siaw-Fong Chung and Chu-Ren Huang</i>	
Inter-clausal Anaphora in Chinese Conditionals: a Multi-factorial Analysis	140
<i>Shunting Chen, Pascal Amsili and Yiming Liang</i>	
A multidimensional analysis: speaking style of learner speech across proficiency levels	149
<i>Tingting Cheng</i>	
Variation across Everyday Conversations: Factor Analysis of Conversations using Semantic Cate- gories of Functional Expressions	160
<i>Yuya Chiba and Ryuichiro Higashinaka</i>	

Dependency Enhanced Contextual Representations for Japanese Temporal Relation Classification <i>Chenjing Geng, Fei Cheng, Masayuki Asahara, Lis Kanashiro Pereira and Ichiro Kobayashi</i>	170
Applying Masked Language Models to Search for Suitable Verbs Used in Academic Writing . . . <i>Chooi Ling Goh</i>	180
Using Distributional Principles for the Semantic Study of Contextual Language Models <i>Olivier Ferret</i>	189
A Comparative Study of Collocation Extraction Methods from the Perspectives of Vocabulary and Grammar: A Case Study in the Field of Journalism <i>Lulu Gu, Yue Pan and Pengyuan Liu</i>	201
A Corpus-based Lexical Semantic Study of Mandarin Verbs of “Tui”and “La” <i>Sun Han and Huang Chu-Ren</i>	211
Unpredictable Attributes in Market Comment Generation <i>Yumi Hamazono, Tatsuya Ishigaki, Yusuke Miyao, Hiroya Takamura and Ichiro Kobayashi</i>	217
Grammatical Error Correction via Supervised Attention in the Vicinity of Errors <i>Hiromichi Ishii, Akihiro Tamura and Takashi Ninomiya</i>	227
Can Monolingual Pre-trained Encoder-Decoder Improve NMT for Distant Language Pairs? <i>Hwichan Kim and Mamoru Komachi</i>	235
Language change in Chinese political discourse based on the relationship between sentence and clause <i>Renkui Hou, Chu-Ren Huang and Kathleen Ahrens</i>	244
Various Errors Improve Neural Grammatical Error Correction <i>Shota Koyama, Hiroya Takamura and Naoaki Okazaki</i>	251
Construction of MBTI Personality Estimation Model Considering Emotional Information <i>Ryota Kishima, Kazuyuki Matsumoto, Minoru Yoshida and Kenji Kita</i>	262
Analyzing Semantic Changes in Japanese Words Using BERT <i>Kazuma Kobayashi, Taichi Aida and Mamoru Komachi</i>	270
Automatic Analysis of Linguistic Features in Journal Articles of Different Academic Impacts with Feature Engineering Techniques <i>Siyu Lei, Ruiying Yang and Chu-Ren Huang</i>	281
HPSG/MRS-Based Natural Language Generation Using Transformer <i>Gyu-Min Lee, Sanghoun Song</i>	291
Improving Adversarial Text Generation with n-Gram Matching <i>Shijie Li and Massimo Piccardi</i>	301
Spatial-temporal attributes in verbal semantics: A corpus-based lexical semantic study of discrim- inating Mandarin near synonyms of “tui1”and “la1” <i>Qiangmei Liang and Chu-Ren Huang</i>	310
Span Detection for Vietnamese Aspect-Based Sentiment Analysis <i>Kim Nguyen Thi Thanh, Sieu Huynh Khai, Phuc Pham Huynh, Luong Phan Luc, Duc-Vu Nguyen and Kiet Nguyen Van</i>	318
The Paradox of Subject-Oriented Resultative Compounds in Taiwan Southern Min	329

Huei-Ling Lin

Joint Chinese Word Segmentation and Part-of-speech Tagging via Two-stage Span Labeling . . . 336
Duc-Vu Nguyen, Linh-Bao Vo, Ngoc-Linh Tran, Kiet Nguyen and Ngan Nguyen

Machine Learning Approach for Depression Detection in Japanese . . . 346
Yuka Niimi, Yutaka Miyaji

KC4Align: Improving Sentence Alignment method for Low-resource Language Pairs . . . 354
Ha Nguyen Tien, Dat Nguyen Huu, Huong Le Thanh, Vinh Nguyen Van and Minh Nguyen Quang

Native Language Identification and Reconstruction of Native Language Relationship Using Japanese
Learner Corpus . . . 364
Mitsuhiro Nishijima and Ying Liu

ALICE++: Adversarial Training for Robust and Effective Temporal Reasoning . . . 373
Lis Pereira, Fei Cheng, Masayuki Asahara and Ichiro Kobayashi

Unknown Intent Detection using Multi-Objective Optimization on Deep Learning Classifiers . . . 383
Prerna Prem, Zishan Ahmad, Asif Ekbal, Shubhashis Sengupta, Sakshi Jain and Roshini Rammani

Schizophrenia Through the Lens of Chinese Classifier: A Preliminary Study . . . 392
Edward Ruoyang Shi

Natural Language Inference using Neural Network and Tableau Method . . . 402
Ayahito Saji, Daiki Takao, Yoshihide Kato and Shigeki Matsubara

Talking with the Theorem Prover to Interactively Solve Natural Language Inference . . . 411
Atsushi Sumita, Yusuke Miyao and Koji Mineshima

Knowledge Grounded Multimodal Dialog Generation in Task-oriented Settings . . . 421
Deeksha Varshney and Anushkha Singh, Asif Ekbal

Study of Near Synonymous Mental-State Verbs: A MARVS Perspective . . . 432
Yuan Tao and Yuan Tao

Deep Speaker Verification Model for Low-Resource Languages and Vietnamese Dataset . . . 442
Dat Vi Thanh, Thanh Pham Viet and Trang Nguyen Thi Thu

Understanding Advanced Level Academic Writing on Syntactic Complexity . . . 452
Miao Wang and Wander Lowie

Investigating effects of gender identity on interpreters' styles: A case study of multidimensional
analysis . . . 463
Xiaoshu Yuan, Yunzhang Shi and Hongyu Wan

A quantitative investigation of English adnominal modifiers . . . 476
Tsy Yih and Haitao Liu

Poster Papers

A MARVS Analysis of Two Chinese Near-Synonymous Verbs of Jumping Based on Chinese Corpora 483
Yingying Peng

Learning Sense-Specific Static Embeddings using Contextualised Word Embeddings as a Proxy . 493
Yi Zhou, Danushka Bollegala

Readability and Linearity . . . 503

<i>Yo Ehara</i>	
Covering a sentence in form and meaning with fewer retrieved sentences	513
<i>Yuan Liu and Yves Lepage</i>	
Stronger Baseline for Robust Results in Multimodal Sentiment Analysis	523
<i>Wei Yang and Jun Ogata</i>	
ParaCotta: Synthetic Multilingual Paraphrase Corpora from the Most Diverse Translation Sample Pair	533
<i>Alham Fikri Aji, Tirana Noor Fatyanosa, Radityo Eko Prasajo, Philip Arthur, Suci Fitriany, Salma Qonitah, Nadhifa Zulfa, Tomi Santoso and Mahendra Data</i>	
Holistic interpretation in locative alternation – Evidence from self-paced reading	543
<i>Rui Cao</i>	
Multi-tasking Dialogue Comprehension with Discourse Parsing	551
<i>Yuchen He, Zhuosheng Zhang and Hai Zhao</i>	
A Corpus-based Gender Study of Hedges in Spoken British English	562
<i>Ruofan Du</i>	
The Annotation of Antonym Information in the ‘Word List by Semantic Principles’	572
<i>Sachi Kato, Masayuki Asahara, Nanami Moriyama and Asami Ogiwara, Makoto Yamazaki</i>	
Incorporating Semantic Textual Similarity and Lexical Matching for Information Retrieval	582
<i>Hiroki Iida and Naoaki Okazaki</i>	
Supervised Word Sense Disambiguation on Taiwan Hakka Polysemy with Neural Network Models: A Case Study of BUN, TUNG and LAU	592
<i>Huei-Ling Lai, Hsiao-Ling Hsu, Jyi-Shane Liu, Chia-Hung Lin and Yanhong Chen</i>	
A Sentiment Analysis of Men’s and Women’s Speech in the BNC64	603
<i>Yong-Hun Lee and Ji-Hye Kim</i>	
Optimizing Domain Specificity of Transformer-based Language Models for Extractive Summariza- tion of Financial News Articles in Korean	611
<i>Huije Lee, Wonsuk Yang, Chaehun Park, Hoyun Song, Eugene Jang and Jong C. Park</i>	
Evaluating a How-to Tip Machine Comprehension Model with QA Examples collected from a Community QA Site	622
<i>Tingxuan Li, Shuting Bai and Fuzhu Zhu, Takehito Utsuro</i>	
Syntactic Distribution of the Semantic Classes of Dative Verbs in English and Cantonese: A Crosslinguistic Perspective	628
<i>Ziying Li and Hanjung Lee</i>	
A Corpus-based Analysis of Prosodic Pauses in <i>bǎ</i> , <i>gěi</i> and <i>ràng</i> Constructions in Taiwan Mandarin	640
<i>Chihkai Lin</i>	
A Corpus-based Lexical Semantic Study of Mandarin Verbs of <i>zhidao</i> and <i>liaojie</i>	646
<i>Yi Liu</i>	
Three Types of Average Dependency Distances of Sentences in a Multilingual Parallel Corpus . . .	652
<i>Masanori Oya</i>	
PS-GAN: Feature augmented text generation in Telugu	662

<i>Pratyusha Musunuru and Varshit Battu</i>	
Whose change of state do resultative adverbialised adjectives describe in Japanese locative-alternating constructions?	672
<i>Masataka Ogawa and Takeshi Kishiyama</i>	
ViVQA: Vietnamese Visual Question Answering	683
<i>Khanh Quoc Tran, An Trong Nguyen, An Tran-Hoai Le and Kiet Van Nguyen</i>	
Monolingual vs multilingual BERTology for Vietnamese extractive multi-document summarization	692
<i>Huy To Quoc, Kiet Nguyen Van, Ngan Nguyen Luu-Thuy and Anh Nguyen Gia-Tuan</i>	
Animosity and suffering: Metaphors of BITTERNESS in English and Chinese	700
<i>Gabor Parti, Andreas Liesenfeld and Chu-Ren Huang</i>	
Analysis of Text-Semantics via Efficient Word Embedding using Variational Mode Decomposition	711
<i>Rohith Ramakrishnan, Anirudh Vadakedath and U Vamsi Krishna, Premjith B and Kp Soman</i>	
Science Mapping of Publications in Natural Language Processing in the Philippines: 2006 to 2020	721
<i>Rachel Edita O. Roxas, Joseph Marvin R. Imperial and Angelica H. De La Cruz</i>	
Construction and Evaluation of Japanese Sentence-BERT Models	731
<i>Naoki Shibayama and Hiroyuki Shinnou</i>	
Verbal Semantics as Eventive Information: The Case of Verb “制服”and “制約”in Chinese Gigaword Corpus	739
<i>Zhan Wang and Ming Ming Chiu</i>	
Study of Similarity Measures as Features in Classification for Answer Sentence Selection Task in Hindi Question Answering: Language-Specific v/s Other Measures	747
<i>Devika Verma, Ramprasad Joshi, Shubhamkar Joshi and Onkar Susladkar</i>	
A Hybrid PSO-GA for Extractive Text Summarization	757
<i>Bui Thi-Mai-Anh, Thi Thu Trang Nguyen, Tran Thi Dinh and Nguyen Thi Hoai</i>	
From Near-synonyms to Divergent Viewpoint Foci: A Corpus-based MARVS Driven Account of Two Verbs of Attention	767
<i>Ziqian Wang and Chu-Ren Huang</i>	
Metaphor Development in Public Discourse Using an ARIMA Time Series Analysis Approach . .	776
<i>Winnie Huiheng Zeng, Dennis Tay and Kathleen Ahrens</i>	