PACLIC 38 (2024)

Proceedings of the 38th Pacific Asia Conference on Language, Information and Computation

07–09 December, 2024
Tokyo University of Foreign Studies
Tokyo, Japan

Nathaniel Oco, Shirley N. Dita, Ariane Macalinga Borlongan, Jong-Bok Kim (Editors)

Foreword

It is our great pleasure to present the proceedings of the 38th Pacific Asia Conference on Language, Information and Computation, or PACLIC 38 (2024), held in hybrid mode on 7–9 December 2024 at the Tokyo University of Foreign Studies (TUFS), Japan.

In a world where languages and computers shape how we live, work, and connect, PACLIC has remained a vital space for reflecting on their evolving relationship. This year's conference brings together researchers from across the region and around the world to address the challenges and possibilities that arise at the intersection of linguistics and computation. In an age marked by linguistic diversity and rapid technological advancement, these conversations are more important than ever.

The contributions in this proceedings reflect a broad spectrum of inquiry—ranging from theoretical explorations in linguistics to cutting-edge developments in natural language processing. They represent not only disciplinary rigor but also a shared commitment to addressing some of the most pressing questions of our time. As Prof. Kayako Hayashi, President of the University, aptly posed during the welcome remarks: How do we foster understanding in multilingual societies? How can we harness technology in ways that enhance, rather than diminish, our human values? What role should artificial intelligence play in shaping our communicative futures?

We are especially grateful to the Tokyo University of Foreign Studies for hosting this gathering—a fitting venue for deepening our understanding of language in all its complexity. We extend our sincere thanks to the authors, reviewers, organizers, student volunteers, and participants who made PACLIC 38 (2024) possible. Your efforts continue to strengthen this community and advance the collective work of research, reflection, and innovation. We also thank Okgi Kim and Soulkee Park for their assistance in the publication of these proceedings.

We hope that the papers included here not only mark the progress of the field but also spark new conversations and collaborations in the years ahead.

Nathaniel Oco, Shirley Dita, Ariane Macalinga Borlongan, Jong-Bok Kim
PACLIC 38 (2024) Program Committee Chairs
(on behalf of the organizing committee)

Organizing Committee

- Ariane Macalinga Borlongan, Conference Co-Chair
- Shirley N. Dita, Conference Co-Chair
- Jasper Kyle Catapang
- Kenichiro Kurusu
- Nathaniel Oco
- Chisato Oda
- Philip Rentillo
- Kim Tiu Selorio

Student Volunteers

- Hiroki Saito
- Parsa Amini
- Zulkar Galip
- Mikhail Alic Go
- Mia Guo
- Yu Hasegawa
- Yuki Ino
- Yusuke Kondo
- Andromeda Labordo
- Hanana Tomizawa
- Jeremy Wu

Program Committee

- Abien Fred Agarap
- Abigail Marticio
- Aileen Bautista Del Rosario
- Aileen Joan Vicente
- Aireen Arnuco
- Aldrin P. Lee
- Alejandro S. Bernardo
- Alen Mateo S. Munoz
- Alvin R. Malicdem
- Ana Cristina Fortes
- Angelina Aquino
- Annie Mae C. Berowa
- Ariane Borlongan
- Ariel Robert Ponce
- Bianca Trish Adolfo
- Charibeth Cheng
- Charmaine Ponay
- Cheng-Zen Yang
- Chisato Oda
- Christian Sy
- Chutamanee Onsuwan
- Dalos D. Miguel
- Edward Tighe
- Elineth Elizabeth Suarez
- Eric Ortega
- Ethel Ong
- Eusebio L. Mique, Jr.
- Gina Ugalingan
- Huei-Ling Lai
- I-Ping Wan
- Jasper Kyle Catapang
- Jean Malolos
- Jeanne Flores-Purpura
- Jennibelle R. Ella
- Jong-Bok Kim
- Jonna Marie A. Lim
- Joseph Marvin Imperial
- Kaela Madrunio
- Kasumi Arciaga
- Kenichiro Kurusu
- Kenneth Loquinte
- Kristine De Leon
- Kristine Kalaw
- Kuang-Hua Chen
- Leif Romeritch Syliongka
- Maria Regina Justina Estuar
- Marie Claire Duque-Cruz
- Marvin Casalan
- Mary Joy Canon
- Melanie Siegel
- Michael B. Dela Fuente
- Michael Tanangkingsing
- Mico Magtira
- Moses Visperas
- Muhammad Afzaal
- Naonori Nagaya
- Nathaniel Oco
- Nattama Pongpairoj
- Nicanor Guinto
- Nimfa Dimaculangan
- Philip Rentillo
- Ponrudee Netisopakul
- Rachelle Lintao
- Rafael Michael O. Paz
- Ramon Rodriguez
- Reginald Neil Recario
- Ria Sagum
- Richard Rillo
- Robert R. Roxas
- Rodney Jubilado
- Romina Gracia C. Cortez
- Satoru Yokoyama
- Shirley Dita
- Teri An Joy Magpale
- Thomas James Tiam-Lee
- Virach Sornlertlamvanich
- Wilkinson Daniel Wong Gonzales
- Wong Tak-Sum
- Yong-Hun Lee

Plenary and Invited Talks

Enhancing L2 Learner Corpus Design Through AI: A Case Study of the JEFLL Corpus Revision

Yukio Tono

Tokyo University of Foreign Studies

Learner corpus research (LCR) emerged in the 1990s, combining corpus linguistics, second language acquisition, and foreign language teaching (Granger, 1998). Initially, LCR focused on comparing native and advanced learner writing through contrastive interlanguage analysis (CIA), examining overuse/underuse patterns and L1 influence. However, SLA researchers' interest in LCR remained limited due to insufficient early acquisition data and weak theoretical foundations in second language acquisition.

This situation has improved through more rigorous data collection focusing on specific aspects of SLA theories (e.g., morpheme order, tense/aspect, verb argument structure). Additionally, there is growing interest in NLP educational applications, including automatic essay evaluation and automated error identification and correction. In recent years, the advent of generative AI has drastically changed the perspective on these NLP applications in education.

In this talk, I will demonstrate how AI can enhance learner corpus design, incorporating recent research paradigms. I will present a work-in-progress report on the comprehensive revision of my learner corpus, the Japanese EFL Learner (JEFLL) Corpus, using generative AI. Specifically, I will outline the construction of parallel texts (original texts, native speaker proofread texts, and AI-corrected texts) with CEFR-level assessment by both humans and AI, based on the JEFLL Corpus. I will also discuss the extension of the JEFLL Corpus for generating texts with improved or downgraded CEFR levels.

AI in Education: Implications for Language Teacher Education

Ee Ling LOW

National Institute of Education, Nanyang Technological University

Artificial intelligence (AI) is revolutionising various sectors, particularly through the use of large language models (LLMs) like ChatGPT, thereby dramatically changing everyday life and work. The impact on language teacher education has been equally profound and far-reaching. This presentation offers an overview of how language education and teacher professional development have the potential to leverage AI's capabilities to provide unprecedented experiences of personalised learning that may be tailored to individual and group needs. These opportunities include adaptive learning systems, collaborative learning across time zones and geographical locations, automated grading, virtual tutoring, and inclusive accessibility tools such as immediate closed captions. While the advantages of AI in language teaching and learning are evident and exciting, the challenges and concerns AI poses are less frequently discussed. Ethical considerations include data privacy issues and the responsible use of AI models to maintain respect and inclusivity for diverse learners. Another concern is how dominant languages, such as English, may contribute to the extinction of other languages and varieties. This keynote will explore potential ethical applications of AI in language education and teacher training. It will address critical considerations surrounding data privacy, algorithmic bias, and the importance of equitable access to AI resources. Using Singapore as a case study, this keynote will examine AI applications at the institute and national levels in language teacher education and presents a faculty technology-enabled learning framework (FacTEL) that is currently being developed and implemented for teacher educators. Future directions for AI in language teacher education includes the need to focus on creating sustainable, ethical, inclusive and equitable learning environments for a diverse range of learners across the globe.

Non-standard morphosyntactic variation in L2 English varieties world-wide: a corpus-based study

Robert Fuchs

University of Bonn

This talk will explore morphosyntactic variation in the global English language complex, specifically across the fourteen L2 and six L1 varieties of English included in the *Corpus of Global Web-based English*, totalling 1.9 billion words – the largest available corpus of global Englishes that includes formal and informal data. We will focus on two studies in particular – one exploring non-standard morphosyntactic variation and one focussing on the (standard) comparative alternation.

Previous research has identified and studied a wide range of non-standard syntactic constructions in global varieties of English. However, there is currently a lack of large-scale corpus-based evidence on non-standard syntactic variation in English from a global perspective (Kortmann, 2021: 299). The term non-standard here refers to features outside the "core" of the language that represents the object of description in English grammars, including those that might be regarded as "colloquial" or "vernacular", such as *there*-existentials with singular agreement.

Across 34 morphosyntactic features drawn from eWAVE 3.0 (Kortmann et al. 2020), a total of 386,661 non-standard and more than 52 million standard occurrences were analysed in a logistic mixed effects regression model. Register variation was accounted for by means of a co-variate for involved discourse following Biber's (1988) Multidimensional Analysis. Results indicate a relatively low degree of non-standardness across both L1 and L2 varieties. The register dimension of involved discourse is the most important variable governing non-standard variation, followed by differences between world regions, and between varieties at different developmental stages. These results were further confirmed by a hierarchical clustering model and a multidimensional scaling analysis.

Study 2 focusses on the comparative alternation. Although grammatical phenomena like the genitive, dative, and particle placement alternations have recently become more frequently investigated in the World Englishes field (e.g., Heller et al., 2017; Szmrecsanyi & Grafmiller, 2023), the choice between the analytic (*more silly*) and the synthetic (*sillier*) comparative variant has received less attention. Grammatical alternations are constrained by a range of factors that have been shown to largely overlap across varieties of English (e.g., Bernaisch et al., 2014; Szmrecsanyi & Grafmiller, 2023). This analysis explores the comparative alternation across 20 varieties of English, using mixed-effects regression. Results show slight differences between Inner- and Outer-Circle Englishes (Kachru, 1985) as well as variation according to degree of nativization (Schneider, 2007); however, considerable overlap in the constraints on the comparative across varieties suggests that the alternation may be part of the common core of the global English language complex. Together, these two studies shed light on what unites and divides different varieties of English in the English language complex. In terms of methodology, the talk will particularly highlight how large amounts of data can be analyzed in depth with big data methods.

Biber, Douglas. 1988. *Variation across speech and writing.* Cambridge University Press, Cambridge.

Bernaisch, T., Gries, S. Th., & Mukherjee, J. (2014). The dative alternation in South Asian English(es). *English World-Wide, 35*(1), 7–31. https://doi.org/10.1075/eww.35.1.02ber

Davies, M., & Fuchs, R. (2015). Expanding horizons in the study of World Englishes with the 1.9 billion word global web-based English corpus (GloWbE). *English World-Wide, 36*(1), 1–28.

Heller, B., Bernaisch, T., & Gries, S. Th. (2017). Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal, 41*(1), 111–144. https://doi.org/10.1515/icame-2017-0005

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge University Press.

Kortmann, B., Lunkenheimer, K., Ehret, K. 2020. (Eds.), *The electronic world atlas of varieties of English, 3.0*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Kortmann, B. 2021. Syntactic variation in English: A global perspective. In: Aarts, B., McMahon, A., Hinrichs, L. (Eds.), *The handbook of English linguistics*. 2nd Edn. Wiley Blackwell, Hoboken NJ , pp. 301–322.

Schneider, E. W. (2007). *Postcolonial English: Varieties around the world.* Cambridge University Press.

Szmrecsanyi, B., & Grafmiller, J. (2023). *Comparative variation analysis: Grammatical alternations in world Englishes.* Cambridge University Press. https://doi.org/10.1017/9781108863742

Large Language Models and Natural Language Processing

Rachel Edita Roxas

University of the Philippines Los Baños

This study presents a systematic literature review on publications on minority languages in large language models and natural language processing. Using the Bibliometrics approach on Scopus-indexed documents published prior to November 2024, analyses and visualization were conducted. Aside from the surge on the number of publications in recent years, collaboration among countries/territories, and the predominance of the computer science subject area are noticeable. The keyword co-occurrence network revealed the prevalence of keywords related to the field of computer science. Schools of thought identified were: 1) Multilingualism and closely-related languages; 2) Performance Evaluation Approaches, and 3) Cross-lingual approaches. We identified the natural language considered in these studies, NLP tasks, technologies used, and social issues and concerns. Conclusions and recommendations for future work are presented.

The *Oxford English Dictionary* and evolving language technologies

Kate Wild

Danica Salazar

Oxford University Press

The *Oxford English Dictionary* (OED), a large historical dictionary which traces the development of meanings and uses of words used across the English-speaking world, has long been an early adopter of new technologies such as digitization, online publication, and electronic text databases. In this paper we discuss how the OED is responding to recent developments in language technologies, in particular Artificial Intelligence (AI) and corpora.

AI has attracted much attention in lexicography especially since the release of ChatGPT in late 2022. OED editors have been experimenting with AI tools and assessing their capabilities for various tasks, including prioritizing new words and senses, suggesting modernized wording for unrevised definitions, and word-sense disambiguation. We give examples of the outputs of such experiments, with a particular focus on uses for global Englishes. We also demonstrate the ways in which AI could transform the user experience of the OED, with a conversational interface powered by a large language model as an alternative to more complex advanced searches.

A longer-standing aim of the OED has been to develop language corpora for lexicographical and academic research. Having built a very large monitor corpus of 21st-century English which covers all major varieties, the next step is to develop similarly diverse corpora for earlier periods of English. We discuss the historical corpora currently used by OED editors, their limitations, and plans for a new resource, including possible interactions with developments in AI.

Querying and challenging the "Generative AI lexicographer" for lexical information

Vincent B Y Ooi

Department of English, Linguistics and Theatre Studies, National University of Singapore

Using examples primarily from English, this talk examines the capabilities of major LLMs - ChatGPT (4-o, 4-o mini) and Gemini (1.0 Pro, 1.5 Pro) - to act as 'generative AI lexicographers' for producing accurate lexical output in relation to established lexicographic principles and practices. Against the backdrop of current discussions on the subject - notably de Schryver (2023), Lew (2024a, 2024b), McKean and Fitgerald (2023), and Rundell (2023) – we can evaluate the performance of these LLMs in benchmarks valued by both corpus linguists and lexicographers alike, including (but not limited to) the ability to handle lexical priming (Hoey 2005,Ooi 2016), semantic prosody/semantic association (Sinclair 2004, Hoey 2005, Stubbs 2001), coverage (Ooi, 2010), and varieties of English (Ooi 2018, 2021, 2023). Examples of such queries include their ability to contrast *kungfu* (as a core/world English item) and *karoshi* (essentially more restricted to the Japanese context), the semantic prosody of 'involuntariness' and 'unpleasant experience' for the verb *undergo*, the ability to define the nominal *killer litter* in the sense of objects thrown from a high building that can injure passers-by below and being predominantly Singapore English, and the ability to minimise cultural bias for *durian* (a fruit often characterised in British and U.S. dictionaries as one that 'smells like hell but tastes like heaven' and being 'pungent' or 'stinky'). Results suggest varying accuracies and nuances according to the version used, but this does not necessarily adhere to their following respective purposes: GPT-4o ('great for most tasks'), GPT-4o mini ('faster for everyday tasks'), Gemini 1.0 Pro ('for an answer with more context'), and Gemini 1.5 Pro ('for a more guided experience').

Gilles-Maurice de Schryver. 2023. Generative AI and Lexicography: the current state of the art using ChatGPT. *International Journal of Lexicography*, 36(4):355-387. https://doi.org/10.1093/ijl/ecad021/

Michael Hoey. 2005. *Lexical Priming: A New Theory of Words and Language.* Routledge, London, UK.

Robert Lew. 2024a. Dictionaries and lexicography in the AI era. *Humanities & Social Sciences Communications*, 11: 1-8. https://doi.org/10.1057/s41599-024-02889-7/

Robert Lew. 2024b. The impact of generative transformers on lexicography. (Invited plenary, the 17th International Conference of the Asian Association for Lexicography, 14 Sep 2024).

Erin McKean and Will Fitzgerald. 2023. The ROI of AI in Lexicography. In *Proceedings of the 16th International Conference of the Asian Association for Lexicography: (Lexicography, Artificial Intelligence, and Dictionary Users).* Seoul: Yonsei University, pages 10–20. https://asialex.org/pdf/Asialex-Proceedings-2023.pdf/

Michael Rundell. 2023. Automating the creation of dictionaries: are we nearly there?. In *Proceedings of the 16th International Conference of the Asian Association for Lexicography: (Lexicography, Artificial Intelligence, and Dictionary Users).* Seoul: Yonsei University, pages 1–9. https://asialex.org/pdf/Asialex-Proceedings-2023.pdf/

John Sinclair. 2004. (co-edited with Ronald Carter). *Trust the Text: Language, Corpus and Discourse.* Routledge, London, UK.

Michael Stubbs. 2001. *Words and phrases: Corpus Studies and Lexical Semantics.* Blackwell, UK.

Vincent B.Y. Ooi. 2010. English Internet lexicography and online dictionaries. *Lexicographica,* 26: 143-154. https://doi.org/10.1515/9783110223231.2.143/

Vincent B.Y. Ooi. 2016. Lexical priming, dictionaries and Asian users of English. In *Proceedings of the 10th International Conference of the Asian Association for Lexicography: (Advancing Language Teaching with Lexicography and Corpus-building).* De La Salle University, Manila, The Philippines, pages 1-13. https://asialex.org/pdf/Asialex-Proceedings-2016.pdf/

Vincent B.Y. Ooi. 2018. Lexicography and World Englishes. In Low Ee Ling, A. Pakir (eds.) *World Englishes: Rethinking Paradigms.* London: Routledge, pages 165-182.

Vincent B.Y. Ooi, 2021. Issues and prospects for incorporating English use in Japan into the dictionary. *Asian Englishes*, 23(1): 62-78. https://doi.org/10.1080/13488678.2021.1876952/

Vincent B Y Ooi. 2023. Varieties of English and their inclusivity in the NAVER Dictionary. In *Proceedings of the 16th International Conference of the Asian Association for Lexicography: (Lexicography, Artificial Intelligence, and Dictionary Users).* Seoul: Yonsei University, pages 134-142. https://asialex.org/pdf/Asialex-Proceedings-2023.pdf/