

Simposium Penyelidikan Korpus Malaysia 2022

Malaysian Corpus Research Symposium

BUKU PROGRAM
PROGRAMME BOOK

28 & 29 SEPTEMBER 2022

BANGI-RESORT HOTEL

Penerokaan/ Penerapan Linguistik Korpus untuk Kemanusiaan Digital/ *Corpus
Linguistics for Digital Humanities*

<https://mcrcs2022.weebly.com>



GALE



MCRN

Malaysian Corpus Research Network





Kandungan *Content*

- 01.** Kata Alu-aluan Dekan
Welcome Message from the Dean

- 02.** Kata Alu-aluan an Pengerusi Pusat Kajian Bahasa & Linguistik
Welcome Message from the Chair Center for Research in Language & Linguistics

- 03.** Kata Alu-aluan Pengerusi Bersama Simposium
Welcome Message from Co-convenors

- 04.** Pengenalan / *Introduction*

- 05.** Jadual Simposium /
Symposium Schedule

- 06.** Bengkel Pasca Simposium /
Post-symposium Workshop

- 07.** Abstrak / *Abstracts*

- 08.** JK Penganjur/ *Organising Committee*



KATA-KATA ALUAN DEKAN *WELCOME MESSAGE FROM THE DEAN*

Syukur ke hadrat Allah SWT kerana dengan limpah kurnia-Nya dapat kita bersama-sama dalam Simposium Penyelidikan Korpus Malaysia 2022. Bagi pihak Fakulti saya mengucapkan selamat bersidang kepada semua pembentang dan peserta simposium kali ini yang dianjurkan oleh Kumpulan Penyelidikan Kemanusiaan Digital, Pusat Kajian Bahasa & Linguistik salah satu Pusat Kajian terunggul di Fakulti Sains Sosial & Kemanusiaan, Universiti Kebangsaan Malaysia.

Simposium ini menyediakan wadah bagi interaksi, eksplorasi, kajian dan pemahaman berkaitan isu-isu penerokaan kajian korpus linguistik yang berkaitan dengan kemanusiaan digital dan data raya. Saya amat berbangga Simposium kai ini dijayakan bersama rakan strategik Malaysian Corpus Research Network, Dewan Bahasa & Pustaka, serta Gale Cengage. Saya berharap gandingan kerjasama ini akan memberi manfaat yang besar kepada semua penyelidik korpus di negara kita. Kerjasama ini selaras dengan naratif baharu UKM yang memaparkan nilai TERAS JIWA (SOUL) dan elemen kebitaraan PRIHATIN menjadi usaha UKM sebagai universiti watan untuk mencapai hala tuju universiti dalam meningkatkan pencapaian ke tahap yang tertinggi.

"FSSK - Memupuk Pemikiran Kritis dan Kreatif"

Selamat bersimposium.



Prof. Dr. Kadaruddin Aiyub
Dekan
Fakulti Sains Sosial dan Kemanusiaan



MCRS2022

KATA ALU-ALUAN PENGERUSI PUSAT KAJIAN BAHASA & LINGUISTIK WELCOME MESSAGE CHAIR CENTER FOR RESEARCH IN LANGUAGE & LINGUISTICS

Assalamualaikum Warahmatullahi Wabarakatuh dan Selamat Sejahtera. Saya mengucapkan setinggi-tinggi tahniah atas kejayaan JK Kerja Simposium Penyelidikan Korpus menganjurkan Simposium 2022 yang bertemakan “Penerokaan/ Penerapan Linguistik Korpus untuk Kemanusiaan Digital/ Corpus Linguistics for Digital Humanities”. Simposium ini bertepatan dengan kemajuan bidang teknologi digital dan penggunaannya yang semakin meluas. Dalam dunia digital, penyelidik bahasa dan sains sosial dituntut untuk menyelami data digital yang seringkali amat besar, rumit lagi mencabar. Melalui Seminar Penyelidikan Korpus kali ini adalah diharapkan para penyelidik, pembentang dan peserta dapat membicarakan cabaran dan tuntutan semasa dalam kajian korpus digital dalam usaha memahami serta mencari penyelesaian kepada pelbagai permasalahan sosial, budaya dan masyarakat.

Kita harus menjadikan seminar ini sebagai wadah untuk memperoleh pelbagai maklumat terkini dalam penyelidikan linguistik korpus. Saya yakin Seminar Penyelidikan Korpus Malaysia 2022 ini dapat dilaksanakan dengan jayanya dan seterusnya menjadi pemangkin kepada para penyelidik korpus. Semoga Allah S.W.T memberikan keberkatan, taufik dan hidayahNya kepada semua pihak yang telah berusaha menjayakan Seminar Penyelidikan Korpus Malaysia 2022 anjuran Pusat Kajian Bahasa & Linguistik, Fakulti Sains Sosial dan Kemanusiaan. Selamat Berseminar.

"FSSK - Memupuk Pemikiran Kritis dan Kreatif"

Prof. Madya Dr. Nor Fariza Mohd Nor
Pengerusi
Pusat Kajian Bahasa & Linguistik





KATA ALU-ALUAN PENGERUSI SIMPOSIUM WELCOME MESSAGE SYMPOSIUM CONVENORS

Selamat datang ke Simposium Penyelidikan Korpus Malaysia 2022. Simposium ini dengan tema Penerokaan/ Penerapan Linguistik Korpus untuk Kemanusiaan Digital/ Corpus Linguistics for Digital Humanities memberi tumpuan terhadap penyelidikan mengenai kemanusiaan digital dan sains data dalam pelbagai bidang, terutamanya bidang Linguistik Korpus.

Minat terhadap penyelidikan empirikal berasaskan dan berpandukan korpus semakin meningkat di Malaysia. Menyedari perkembangan ini, sekumpulan penyelidik korpus telah berbincang untuk menganjurkan persidangan watan tentang kajian korpus. Maka lahirlah Simposium Penyelidikan Korpus Malaysia yang bermula di USM pada tahun 2012 sebagai Malaysia Corpus Research Colloquium dan seterusnya telah dianjurkan oleh (UPM) 2015, (UTM) 2018 (UiTM) 2020 di bawah tajuk Malaysia Corpus Research Symposium.

Bagi pihak penganjur, kami mengucapkan terimakasih kepada Associate Professor Dr Michael Barlow atas kesudian menyampaikan ucaptama kali ini dan juga kepada Prof Paul Baker sebagai pembentang kertas kerja jemputan khas serta semua pembentang kertas kerja. Kita percaya simposium ini akan mencapai objektif menjadi wadah untuk pertukaran ide serta membina rangkaian antara penyelidik berasas / berpandu korpus.

Akhir kata, terima kasih kepada rakan kerjasama MCRN, DBP dan Gale atas sokongan padu menjayakan simposium ini. Sekalung penghargaan kepada pihak Dekan dan Pengerusi Pusat Kajian atas kepercayaan dan sokongan. Juga kepada semua sukarelawan dan ahli jawatankuasa, kami ucapkan terima kasih menjayakan simposium ini.



**Prof Dato Dr. Imran Ho Abdullah
& Dr Azhar Jaludin
Pengerusi Bersama
Simposium Penyelidikan Korpus
Malaysia 2022**



OBJEKTIF SIMPOSIUM

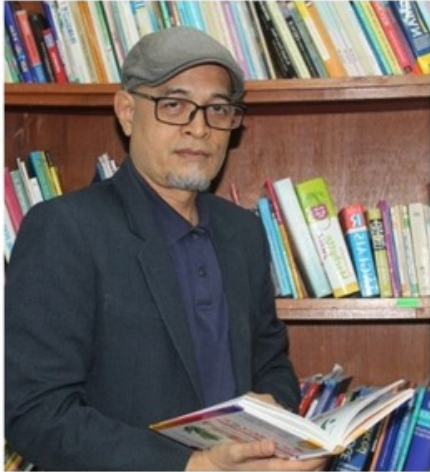
Sarjana tersohor dalam bidang Linguistik Korpus umum bersetuju bahawa peranan penyelidikan korpus serta aplikasi korpus dalam era digital akan semakin menonjol. Ini seiring dengan penekanan bukan sahaja kepada data raya dalam bentuk digital tapi juga selari dengan pembentukan budaya kemanusiaan dan masyarakat digital. Satu aspek budaya digital adalah bahasa yang digunakan untuk pelbagai aktiviti harian dari komunikasi di media sosial, transaksi dan carian atas talian, Dengan kata lain segala kegiatan masyarakat sama ada ilmiah, sosial, politik, kerohanian, ekonomi, kekeluargaan, kesihatan yang melibatkan penggunaan bahasa dan komunikasi akan wujud dalam bentuk digital yang seterusnya boleh dikumpulkan menjadi korpus untuk penyelidikan.

Atas perkembangan ini, simposium ini diadakan untuk

1. Menghimpunkan ahli akademik dan penyelidikan tempatan dan nusantara dalam bidang Linguistik Korpus dan Kemanusiaan Digital.
2. Memberi tumpuan terhadap isu-isu yang timbul daripada penyelidikan terkini dalam bidang Linguistik Korpus dan Kemanusiaan Digital
3. Berkongsi pengalaman dalam penyelidikan baru dan dapatan kajian menerusi interaksi dan penerbitan akademik



BENGGKEL PASCA SIMPOSIUM



Dr. Azhar Jaludin
Centre for Research in Language and Linguistics
Universiti Kebangsaan Malaysia

WORKSHOP
Data Mining Part II: Web
Crawling for Language Data

- * Learn how to compile data for corpus development with hands-on on searching and gathering data using available packages

Post-symposium Workshop

**2022 MALAYSIAN CORPUS
RESEARCH SYMPOSIUM**



MALAYSIAN CORPUS AND
APPLIED LINGUISTICS
MONTHLY

WEBINAR

3.00 TO 5.00 P.M.

(MYT : UTC +8)

SEPTEMBER 30TH, 2022



Moderator
Anis Nadiah Che Abdul Rahman
Universiti Kebangsaan Malaysia

#MCRNWebinar



JADUAL / SCHEDULE

Rabu, 28 September 2022

0800	Pendaftaran / <i>Registration (kudapan / morning coffee)</i>
0845	Majlis Perasmian / <i>Opening Ceremony</i> Negaraku Bacaan Doa / <i>Doa Recital</i> Alu-aluan Dekan / <i>Welcome Speech by the Dean</i>
0920-1020	UCAPTAMA / KEYNOTE Using corpus techniques in social media research. Assoc. Prof Michael Barlow (University of Auckland) *(8:20 pm Austin Texas USA 27 September) online
1020 - 1140	Sesi/Session 1 1. Bahasa Tidak Santun dalam Penulisan Blog politik di Malaysia : Analisis Korpus. Azizul Ismail (DBP) 2. Prosodi Semantik 'Wanita' dalam Korpus Hansard Malaysia. Anis Nadiah Che Abdul Rahman (UKM) 3. Pemilihan kata di Facebook kesan emosi dan tekanan semasa perintah Kawalan Pergerakan (PKP). Hishamudin bin Isam (UUM) 4. Idiolek Pengarang: Shahnnon ialah Shahnnon dan Arena Wati bukan Abdullah Hussain. Norhafizah Mohamed Husin (DBP)
1140 - 1245	PLENARI GALE 1 / GALE PLENARY 1 1. Does Opinion Matter? A Sentiment Analysis of Opinion News on Climate Change in News Discourse. Nor Fariza Mohd Nor (UKM) 2. Corpus Linguistics Research in Indonesia. Prihantoro (Universitas Diponegoro)
1245 - 1400	Makan Tengahari / <i>Lunch break</i>
1400 - 1530	Sesi/Session 2 (dalam talian / online) 1. Making assumption about light verb constructions in Malaysian English. Christina Ong Sook Beng (UTAR/USM) 2. Investigating the multiword units presented in the English textbooks for Vietnamese EFL tertiary-level students. Hien Hoang (University of Queensland) 3. Collostructional Analysis on future 'will' in Malaysian English Online News. Deborah Tirtania & Azhar Jaludin (UKM) 4. A Multi-dimensional Analysis of Risk Management Reports by Malaysian and American Banks. Khairul Firhan Yusob (UiTM), Afida Mohamad Ali (UPM), Mei Yuit Chan (UPM) & Geok Imm Lee (UPM) 5. The Design and Development of MUET Corpus. Noorli Khamis (UTeM) 6. A Corpus Analysis of Metaphorical Expressions in Covid-19 Open Research Dataset (CORD19) - Shazia Riaz Dar (GIFT U)
1530-1600	Kudapan / Evening tea
1600 - 1700	PENCERAMAH JEMPUTAN / GUEST SPEAKER Sex differences on a forum about anxiety. Prof. Paul Baker (University of Lancaster) 28th September 2022 (0900 UK Time)



0800 – 0900	Kudapan/ <i>Morning coffee</i>
0900 – 1020	<p>Sesi/Session 3</p> <ol style="list-style-type: none"> 1. A Corpus-Assisted Critical Discourse Analysis of Malaysian Religious COVID-19 Anti-Vaccine Telegram Posts. Mohd Nazriq Noor Ahmad (UM) 2. A Corpus-Driven Analysis of Lexical Variation and Change in Malaysian Twitter. Noraishah Gulnazir (UKM) 3. Deconstructing News Media Stereotypes: An Analysis of Western News Discourse on The Palm Oil Industry in Southeast Asia Ravichandran Vengadasamy (UKM) & Mohamad Iadid Ashrai Hassannudin (UiTM) 4. Epistemic Modal Verbs in George W. Bush's Presidential Speeches: Persuading Americans to go to war in Iraq. Yasir Ahmed Mohammed Ridha Al-Sabbagh (UKM)
1020 - 1220	<p>PLENARI GALE 2 / GALE PLENARY 2</p> <ol style="list-style-type: none"> 1. Corpus Stylistics: Integrating English Language, Literature and Translation in the Digital Age. Raksangob Wijitsopon, Ph.D (10.am Bangkok, Thailand 29th September) - online 2. Research Trends in Corpus Linguistics. A Bibliometric Analysis of Two Decades of Scopus-indexed Corpus Linguistics Research in Arts and Humanities - Peter Crosthwaite (University of Queensland) – online 3. Language Variety Across Time: A Case Study of the Diachronic Malaysia English Corpus (DMEC). Siti Aeisha Joharry (UiTM) 4. The linkage between Chief executive officers statements and Corporate Sustainability reporting: an examination of visuals and verbal texts. Asia Munir (UTP) , Shahrina Md Nordin (UTP) , Ruzy Suliza Hashim (UKM) - online
1230 – 1400	Makan Tengahari / Lunch break
1400 – 1445	<p>Sesi / Session 4</p> <ol style="list-style-type: none"> 1. Tanggapan Penutur Asing Dewasa terhadap E-buku Interaktif berdasarkan Pendekatan Model Penerimaan Teknologi (TAM) Marene Mohammad (UTM) 2. Preferensi semantik: Analisis berasaskan korpus tentang Perubahan Iklim dalam Korpus Hansard Malaysia (MHC). Muhammad Zakwan Mohd Izam (IIUM/UKM) 3. Peranan Kemanusiaan Digital ke Arah Perkembangan Penyelidikan Proses Penterjemahan Digital (DTPR) Anis Shahirah Abdul Sukur (USM) – online
1445 - 1515	<p>SESI KORPUS DBP / DBP CORPUS SESSION</p> <p>Mohd Fadzli bin Tajuid, Ketua Bahagian Peristilahan & Leksikologi, Dewan Bahasa & Pustaka Pelancaran Glosari Linguistik Korpus (DBP)</p>
1515 - 1615	<p>FORUM MCRN</p> <p>Ke Arah Pemeraksanaan Penyelidikan Korpus di Malaysia</p> <ol style="list-style-type: none"> 1. Prof Emeritus Dato Dr Tengku Mohd Tengku Sembok (UPNM) 2. Prof Dr Hajar Abdul Rahim (USM/ MCRN) 3. Dr Rusmadi Baharuddin (DBP) 4. Prof Madya Dr. Afida Mohamad Ali (UPM/MCRN) <p>Majlis Penutup & Sesi Bergambar / <i>Closing Ceremony</i></p>
1630	Kudapan / Evening break



ABSTRAK/ ABSTRACT UCAPTAMA / KEYNOTE

Using corpus techniques in social media research.

Michael Barlow

University of Auckland, New Zealand

In this talk I discuss a variety of techniques related to the collection and analysis of social media data through a case study of TripAdvisor hotel reviews. In the study, I look for differences in the language used and in the topics covered and explore how those differences might be linked with gender. I will first demonstrate how to collect data from the TripAdvisor website to form a corpus, then discuss some general issues and techniques related to the analysis of social media data using typical corpus techniques such as searches, semantic tagging, and keyword analysis.

In particular, I investigate three questions:

- Are there gender differences in the language used by commenters in TripAdvisor posts?
- Do commenters tend to talk about different things in TripAdvisor posts?
- Are there gender differences in identity work?

I downloaded four types of data from the posts on the website: the username, the location, the heading, and the text. Using the usernames (with care), we can create corpora based on the male and female contributions and then carry out a keyword analysis. We find, for example, that the top keywords in the header for women are: lovely, find, spacious, way, and luxurious. For males, the top keywords in the header are downtown, four, property, views and exceptional. We can also perform a keyword analysis using semantic tags and this provides some insights into categories rather than words. For instance, the category “Information technology and computing” was more prominent in the male posts. It has been reported that women are more likely to overtly signal their gender and this was true in the construction of usernames. There were more instances of names such as hikergirl1 than names that include boy, as in travelboy60. Names with guy, were, however, quite popular: guy4travel, afineguy, etc. I also carried out a brute force method for determining the topics covered in the posts and I found that, in general, there were no gender differences with respect to the content of the posts. For example, men and women talked about food/drink to more or less the same extent. Here, I have given a brief description of the study. In the presentation, I also discuss some techniques and issues related to research on social media.

BIODATA:

Michael Barlow received his PhD in Linguistics from Stanford University. He is currently Associate Professor in the Applied Linguistics and Language Teaching Department at the University of Auckland in New Zealand where he teaches courses on Corpus Studies in Applied Linguistics and Computer-Assisted Language Learning. He has created several text analysis programs including two concordancers, MonoConc and ParaConc, and a collocation extraction program called Collocate. A new program, WordSkew, facilitates the analysis of written and spoken discourse. Dr. Barlow has written books and articles related to the use of corpora in theoretical and applied linguistics and he regularly gives keynotes and workshops at institutions and conferences around the world.

GUEST SPEAKER

Sex differences on a forum about anxiety

Prof. Paul Baker
University of Lancaster

Anxiety is a growing, worldwide phenomenon. The World Health Organization estimates that there are 264 million people living with anxiety and diagnoses of anxiety disorders are more common among women than men (4.6% compared to 2.6% at the global level) (World Health Organization 2017). This book examines 23 million words of text posted to the Anxiety Support forum of the social networking service Health Unlocked between March 2012 and October 2020, comprising 294,082 separate posts.

The way that people conceptualise feelings or conditions is significant. Chan, Chan and Kwok (2015) found that catastrophising was a positive predictor of anxiety among adolescents while Chen, Chen and Yang (2019) have described a study where individuals who were instructed to anthropomorphise sadness or happiness reported less experience of that emotion afterwards.

A small number of posts were written by people who did not identify as male or female and these posts were read in order to identify patterns or discourses relating to gender and sex identities. Then, lexical differences between male and female posters were retrieved by comparing the two sets of posts against one another to identify keywords and key multi word terms, using the corpus analysis tool Sketch Engine. These were categorised by hand and concordances examined to identify 1) how posters characterised anxiety and their relationship to it 2) strategies for managing anxiety 3) reasons for anxiety 4) interactional style on the forum. Additionally, words which explicitly referenced identities (e.g. female, man etc) were also subjected to concordance analyses.

The analysis shows trends which indicate that the ways that individuals understand and write about anxiety are filtered through gendered discourses. The findings have implications for the ways that people with anxiety and health practitioners attempt to understand and resolve anxiety.



GALE PLENARY 1

Does Opinion Matter? A Sentiment Analysis of Opinion News on Climate Change in News Discourse.

Nor Fariza Mohd Nor

Universiti Kebangsaan Malaysia

The growing threat of climate change is driven largely by human activity and social behaviours. In the context of climate change, public opinion matters because opinion from the public can either shape government's actions or inactions on climate change issue. Past studies have documented public opinion on climate change using survey method. The key contribution of this study is understanding how media coverage of climate change relates to public opinion, specifically Malaysian public opinion, which has not been explored in the Southeast Asia region. This study is guided by corpus linguistics approach and the three key dimensions of climate change opinion framework (belief and concern, salience and support for government action); to identify adjective sentiment lexicons in opinion news, to determine the categorisation of the adjective sentiment lexicons and examine Malaysian public opinion on climate change based on the three key dimensions of climate change opinion. The results revealed a high percentage of negative adjective sentiment, with majority of the adjective sentiments in the salience dimension. Overall, this study demonstrates that the public are aware and are concern of the growing threat of climate change, thus, decision makers should work together with the public to find ways to reduce climate change effect and find solutions to the climate change crisis.

Biodata:

Nor Fariza Mohd Nor, is an Associate Professor at the Center of Research for Language and Linguistics, Faculty of Social Sciences and Humanities, National University of Malaysia (UKM). She received her Ph.D degree in Applied Linguistics from University of Malaya and Master degree in English for Specific Purpose from the University of Warwick, U.K. Her Bachelor (Hons) degree is in Modern English Language and Education, from the University of Lancaster, U.K. Nor Fariza is actively involved in research. Her main research includes work in discourse and critical discourse studies, such as work on the Malaysian Hansard corpus, newspaper and social media corpus. She has recently been working on discourse studies by combining critical approach with topic modelling and word embedding approaches. She is currently involved in research in Digital Humanities and data analytics (social science perspective). She was the Editor-in- Chief for GEMA Online Journal. of Language Studies. The journal is indexed in established databases such as Elsevier (Scopus), ESCI (WoS), DOAJ, Wiley-Blackwell, EBSCO, Linguist List, Asean Citation Index and Islamic World Citation Center. She has published extensively, both in local and international journals, which include ISI, Scopus, referred and indexed journals in her area of expertise. Nor Fariza is a member of Malaysian Applied Linguistics Association and GLoCALL.



GALE PLENARY 1

Corpus Linguistics Research in Indonesia.

Prihantoro

Universitas Diponegoro

I here present a macro view of Corpus Linguistics (CL) research in Indonesia, which I divide into three categories: corpus data creation, processing and analysis/interpretation. In terms of corpus data creation/compilation, I highlight some existing Indonesian corpora as well as corpora for other languages spoken in Indonesia. I also discuss a number of tools created for the processing of Indonesian language corpora such as Part of Speech (POS) and morphological annotation systems. In terms of corpus data analysis, I present the results of my inspection on recent linguistic conferences in Indonesia. The inspection shows that studies in which corpus linguistics techniques are applied vary in terms of linguistics disciplines or target languages. A number of institutions greatly contribute to the development of CL research such as MLI (Masyarakat Linguistik Indonesia/the Linguistics Society of Indonesia), Badan Pengembangan dan Pembinaan Bahasa Indonesia (Language Development and Fostering Agency of Indonesia), INACL (Indonesian Association of Computational Linguistics), and various research centres or study programs from universities in Indonesia. Some challenges (e.g. creation of quality corpora, balance across languages and disciplines, high degree of language diversity, finance and human resources) are indeed present, but I believe that collaboration across institutions and disciplines is key to the development of CL research in Indonesia.

Biodata:

Dr. Prihantoro earned his PhD from Lancaster University and is affiliated with Universitas Diponegoro, Indonesia. He manages some corpora in CQPweb Lancaster. He is the author of SANTI-morf (a morphological annotation system) and Buku Referensi Pengantar Linguistik Korpus (Introduction to corpus linguistics book, written in Indonesian). Reach him via prihantoro@live.undip.ac.id.



PLENARI GALE 2 / GALE PLENARY 2

Corpus Stylistics: Integrating English Language, Literature and Translation in the Digital Age.

Raksangob Wijitsopon

Chulalongkorn University

Biodata:

Raksangob Wijitsopon is an associate professor in English at Department of English, Faculty of Arts, Chulalongkorn University. Currently, she is also Deputy Director of the English as an International Language Program (EIL) (MA/ PhD) (Interdisciplinary & International Program). She is particularly interested in using a corpus linguistic approach to analysis of texts of various kinds, ranging from prose fiction (especially classic works), news, magazine columns, learner essays, to translations, etc. She is especially keen on the English language-literature interface, as well as connecting studies of English language and literature to the real world for English major studies.

Language Variety Across Time: A Case Study of the Diachronic Malaysia English Corpus (DMEC)

Siti Aeisha Joharry

UiTM

Building a diachronic corpus for language analysis is a big feat and therefore not surprising that these types of corpora are little in number. In this talk, I will share interesting findings from the Diachronic Malaysian English Corpus (DMEC) to showcase how studies in digital humanities can benefit from the field of corpus linguistics. Using a number of classic corpus techniques, I discuss how analysing language in this day and age can be made more visual as well as empirical for researchers in the social sciences and humanities

Biodata:

Siti Aeisha Joharry holds a PhD in linguistics from The University of Sydney, Australia. In 2017, she joined Akademi Pengajian Bahasa, UiTM Shah Alam, Malaysia where she is currently the programme coordinator (Penyelaras) for the English for Professional Communication Bachelors Programme and resource person for a postgraduate course on corpus linguistics. She is part of the steering committee for the Malaysian Corpus Research Network (MCRN) that hosts monthly webinars on corpus research in Malaysia and also part of the international advisory board for the Institute for Corpus Research, Incheon National University in South Korea. She can be found on both LinkedIn and ResearchGate.



Research Trends in Corpus Linguistics. A Bibliometric Analysis of Two Decades of Scopus-indexed Corpus Linguistics Research in Arts and Humanities

Peter Crosthwaite

University of Queensland

This paper uses a bibliometric analysis to map the field of Corpus Linguistics (CL) research in arts and humanities over the last 20 years, while tracking changes in the popular CL research topics, outlets, influential authors, and geographical origins of published CL research. Based on a collection of the metadata of 5,829 CL-related articles from 429 Scopus-indexed journals, our results reveal an increase in corpus-assisted discourse studies, lexical bundles and academic writing, a reduction (in relative terms) of studies on grammar and translation, and the introduction of new topics including multilingualism and social media. CL studies span 193 total languages/dialects across the period, and we report a significant rise in CL studies in Chinese, Russian, Spanish, and Italian over the past decade. A number of influential CL researchers have remained highly productive over the past two decades, while clusters of CL researchers are identified spanning a range of (inter)disciplinary research areas. Although the USA and the UK still account for the highest raw frequency of Scopus-indexed CL research, their slower relative increase compared with that of China, Poland, South Korea, Japan and many others is evidence that the global reach of CL research has expanded considerably over this 20-year period. Our data reveal links between developments in CL research and diachronic socio-cultural developments in applied linguistics, and society more generally. We discuss the implications of these findings for the field and provide insights into what CL research might come next.



The Linkage between Chief Executive Officers Statements and Corporate Sustainability Reporting: An Examination of Visuals and Verbal Texts

Asia Munir, Shahrina Md Nordin, Ruzy Suliza Hashim
*Universiti Teknologi PETRONAS,
Universiti Kebangsaan Malaysia*

According to research on corporate social responsibility (CSR) reporting, there is a rising lack of transparency in the information given and concerns about overall reporting procedures. One area that requires investigation is how chief executive officers (CEOs) transmit CSR messaging. This article aims to look into the relationship between CEO statements (words and visuals) about CSR and the performativity of such communication.

The study examined CEO statements from five Malaysian oil and gas companies published in standalone sustainability and annual reports in 2016, 2017, 2018 and 2019. Discourse analysis is used to examine the texts and pictures. The data reveal three major discourses (economic, environmental, and social) and additional discourses (achievements, recognition, and challenges). The wording and visuals are deemed to be unclear and inconsistent, leaving stakeholders to draw their conclusions about the presented information. According to the findings, while leaders can be more direct with their stakeholders, the opportunity is not always taken advantage of. Overall, the data indicate growing scientism in CEO messaging on CSR as a tool to improve the perceived accountability and transparency of the business. The report also proposes ways to improve. This research adds to the establishment of several sorts of discourses supported by CEOs in their CSR statements in the Malaysian setting. The identified discourses provide fascinating insights into how leaders perceive CSR.



Bahasa Tidak Santun dalam Penulisan Blog politik di Malaysia : Analisis Korpus.

Azizul Ismail

Dewan Bahasa & Pustaka

Kajian ini bertujuan untuk mengenal pasti dan menganalisis bahasa yang tidak santun berdasarkan maksud leksikal dan analisis korpus DBP dalam blog politik di Malaysia. Penyelidikan ini mengaplikasikan pendekatan Pengelasan Kata Jay, 2000; Pinker, 2008; Wilson, 1998 dengan menggunakan kaedah kajian kepustakaan, soal selidik, dan analisis teks yang terdiri daripada 10 blog politik di Malaysia. Dapatan kajian menunjukkan terdapat 94 perkataan tidak santun yang dibahagikan kepada tujuh kategori dengan 550 kekerapan secara keseluruhannya dalam blog politik yang dikaji. Kesimpulannya, dapatan ini amat penting kepada penulis dan pengikut blog politik, pengkaji bahasa serta masyarakat umum supaya hasil penulisan blog politik mematuhi prinsip kesantunan bahasa.



Prosodi Semantik 'Wanita' dalam Korpus Hansard Malaysia

Anis Nadiah Che Abdul Rahman
Universiti Kebangsaan Malaysia

Wacana yang berkisarkan wanita telah menjadi subjek di dalam pelbagai kajian linguistik. Isu berkaitan wanita juga kian menarik perhatian terutamanya dalam kalangan ahli parlimen. Secara umumnya, penglibatan wanita di parlimen dapat membantu untuk membentuk identiti mereka dan menyumbang kepada dasar dan pelan tindakan kerajaan. Walau bagaimanapun, penyelidikan mengenai pandangan anggota parlimen Malaysia mengenai isu-isu wanita di dalam parlimen adalah terhad. Justeru, peralatan korpus digunakan di dalam penyelidikan ini untuk menganalisis pola pandangan penggubal undang-undang melalui data raya untuk melihat ciri-ciri isu wanita dalam parlimen Malaysia. Kajian ini menggunakan analisis berpandukan korpus untuk meneroka corak perwakilan perkataan 'wanita' dalam korpus Hansard yang bersaiz 164 juta perkataan iaitu perbahasan parlimen Malaysia dari Parlimen 1 sehingga Parlimen 14. Korpus Hansard Malaysia ini dianalisis menggunakan WordSmith Tools 7.0 dengan memanipulasi senarai kekerapan kata, kolokasi kata nama dan kata kunci dalam konteks untuk mengenal pasti pola prosodi semantik perkataan 'wanita.' Kajian ini menggunakan teori prosodi semantik untuk mengkategorikan makna yang berkisarkan perkataan 'wanita'. Dapatan kajian menunjukkan bahawa isu wanita dibincangkan dengan lebih positif di dalam Parlimen Malaysia 1 hingga 14. Menariknya, kolokasi yang kelihatan ketara sifat negatifnya seperti 'pelacuran' dan 'dadah' pula berkemungkinan untuk mempunyai konteks yang positif dalam perbahasan parlimen Malaysia dari perspektif perundangan. Secara keseluruhan, penemuan ini membantu pembinaan satu kerangka untuk memahami isu-isu wanita melalui perwakilan semantiknya dalam perbahasan parlimen Malaysia.



Pemilihan Kata di Facebook. Kesan Emosi dan Tekanan Semasa Perintah Kawalan Pergerakan (PKP)

Hishamudin Isam

Universiti Utara Malaysia

Perubahan hidup yang berlaku semasa Perintah Kawalan Pergerakan (PKP) yang berulang dari 2020 hingga 2021 telah memberi implikasi besar kepada penduduk dunia sehingga menyebabkan masalah gangguan emosi dan tekanan jiwa. Sebagai jalan keluar, ramai dalam kalangan mereka yang terkesan menjadikan laman media sosial sebagai tempat untuk meluahkan emosi demi meredakan tekanan yang dihadapi. Bertitik tolak daripada hal tersebut, kertas kerja ini bertujuan meneliti pemilihan kata, kesan emosi dan tekanan jiwa dalam kalangan pengguna media sosial Facebook (Fb). Dengan memanfaatkan kaedah Operasi Pengiklanan Maya (OPM) sebagai operasi standard untuk memperoleh data daripada pangkalan data raya yang bersifat umum dan terbuka seperti Fb, seramai 120 orang pemilik akaun yang mengaku mengalami gangguan emosi dan tekanan jiwa telah dipilih sebagai responden. Dapatan daripada keseluruhan data yang berjumlah 3,456,162 (token) dan 469,350 (type) menunjukkan, terdapat tiga (3) kata yang paling kerap diulang-ulang penggunaannya dalam status siaran dan komen balas iaitu kata takut dengan jumlah sebanyak 18,792 (frequency), diikuti dengan kata marah (16, 569) dan kata sedih (13,801). Perilaku semantik bagi ketiga-tiga kata tersebut yang dianalisis menggunakan teori prosodi semantik membayangkan penuturnya sedang bergolak dengan gangguan emosi dan tekanan jiwa yang dahsyat, yang jika tidak dirawat segera, boleh mengakibatkan kesan yang lebih parah seperti amuk atau kehilangan nyawa. Kajian-kajian seperti ini wajar digiatkan untuk memastikan kluster bidang kemanusiaan digital pengkhususan linguistik dapat dimanfaatkan untuk membantu bidang-bidang ilmu lain. Hal ini selari dengan matlamat Revolusi Industri 4.0 yang berhasrat untuk menaik taraf semua bidang penyelidikan, agar sinergi ilmu antara bidang dapat direalisasikan dengan jayanya.



Diolk Pengarang: Shahnnon lalah Shahnnon Dan Arena Wati Bukan Abdullah Hussain

Norhafizah Mohamed Husin

Dewan Bahasa dan Pustaka

Kajian ini bertujuan melihat dan mengenal pasti idiolek tiga orang sasterawan negara, iaitu SN Arena Wati, SN Abdullah Hussain dan SN Shahnnon Ahmad. Pemilihan pengarang berdasarkan senarai karya yang ada di dalam Korpus DBP. Sebanyak dua karya bagi setiap seorang SN dianalisis bagi mencari persamaan dan perbezaan dalam gaya penggunaan bahasa mereka masing-masing. Data karya dianalisis secara utuh tanpa sebarang pemurnian. Ciri idiolek dilihat menerusi nisbah type-token, panjang kata, panjang ayat, pola penggunaan tiga kata tugas yang menduduki kelompok lima kata paling kerap digunakan oleh kesemua sasterawan negara yang dipilih, iaitu dan, itu dan yang, dan lima Proses yang paling kerap digunakan dalam karya. Kajian kuantitatif ini menggabungkan pendekatan korpus linguistik dan analisis transitiviti Teori Sistemik-Fungsional oleh Halliday (1981, 1985). Kajian mendapati wujud beberapa ciri idiolek individu yang tekal merentas karya. Kesimpulannya, pendekatan yang dipilih dapat memperlihatkan idiolek unik ketiga-tiga pengarang yang menyerlahkan identiti penulisan dan pola penggunaan bahasa masing-masing. Identiti ini merupakan 'cap jari penulisan' yang boleh dimanfaatkan dalam kajian linguistik forensik Melayu.



Making assumption about light verb constructions in Malaysian English.

Christina Ong Sook Beng

UTAR/USM

This study aims to find out the structural and functional patterns of light verb constructions (LVCs) headed by make in Malaysian English. A typical LVC is made up of a semantically “empty” verb or light verb (e.g. make), an indefinite article and a deverbal noun like example (1).

- (1) We shouldn't make a joke of suicide!
- (2) You are making invalid assumption about me.

To identify potential grammatical innovations, I created a 100-million word general corpus made up of threads from Lowyat.Net, a popular Internet forum in Malaysia and used the British National Corpus (BNC) as the reference corpus. I take a more inclusive approach by considering variants such as zero article and descriptive adjective in example (2) that may not be acceptable by native speakers to be relevant for my analysis. The structural analysis reveals that Malaysians tend to overuse non-isomorphic nouns, zero article LVCs and those taking determiners as well as descriptive adjectives (e.g. invalid in example 2) in the modifier slots. The functional analysis shows that most Malaysian English make LVCs are atelic indicating unbounded actions. It is possibly due to the high frequency of abstract and mass deverbal nouns that can never co-occur with the indefinite article. This finding contradicts hypothesis of related work which suggest that the function of LVCs is to transform aimless actions into achievements. This corpus study thus highlights the flexible nature of LVCs and the resulting patterns of nativisation in Malaysian English. Also, the widely used non-prototypical LVCs should be treated as a sign of endonormativity.



Investigating the multiword units presented in the English textbooks for Vietnamese EFL tertiary-level students

Hien Hoang

University of Queensland

Little is known about multiword unit (MWU) vocabulary presented in English textbooks used in EFL contexts. In recognition of this need, this corpus-based study explores the range of MWUs presented in the reading input and listening input in the English textbooks used within a tertiary-level EFL program in Vietnam. The study explains the procedures of arriving at the list of MWUs in English textbooks, which is then used to compare MWU occurrence rate and frequencies between written and spoken corpora. In terms of procedure, MWUs were automatically identified based on frequency, distributional and statistical score data, before being subject to manual refinement (following Simpson-Vlach & Ellis, 2010) and evaluation for pedagogical validity. A final list of 263 pedagogically valid MWUs was created. The cross-corpus analysis revealed that the overall MWU occurrence rate was significantly greater in the listening input than in the reading input, representing a fairly large effect size. High frequency spoken MWUs were found to have a higher frequency count than high frequency written ones. When the composition of occurrence rate was examined, reading input was found to have used a much larger normalized frequency and also a much wider variety of MWUs than listening input. The study also found that referential expressions are the most commonly used category in both reading input and listening input, whereas discourse organizers seem to be exposed to a lesser extent compared to both referential and stance expressions. Pedagogical implications are discussed regarding the need for placing more emphasis on spoken MWUs in language teaching and for highlighting learners' awareness of written/spoken register variation involving MWUs.



Collostructional Analysis on future 'will' in Malaysian English Online News.

Deborah Tirtania & Azhar Jaludin

Universiti Kebangsaan Malaysia

As the realization of Construction Grammar theory as the basis of collostructional analysis, this study aimed to find out the attracted and repelled verbs occur in the construction of future will of Covid-19 vaccine in Malaysia and how these verbs indicate the predictability of vaccination for locals. A specific context of Covid-19 vaccine in Malaysia from February to April in 721 Malaysian English online newspaper were used as the source of data in this study. A sequential explanatory design was used in this research which consisted of simple collexeme analysis to get a list of attracted and repelled verbs in future will and qualitative descriptive to explain how those verbs indicate the predictability of vaccination for Malaysian. This study proved that aspectual verbs (e.g. start, begin, continue), engender verbs (e.g. benefit, enable, result), and verbs of desire (e.g. need, deliberate) are the top three of attracted verbs classification. Meanwhile, in repelled verbs classification, it is nominated by verbs with predicative complement (e.g. accept, buy, address), verbs of creation and transformation (e.g. raise, assist), and engender verbs as well as aspectual verbs. Moreover, this study also found that during the selected three months of 2021, those attracted verbs indicate the predictability of Covid-19 vaccine in Malaysia is limited to the distribution and how the vaccination program was prepared.



A Multi-dimensional Analysis of Risk Management Reports by Malaysian and American Banks

Khairul Firhan Yusob, Afida Mohamad Ali, Mei Yuit Chan & Geok Imm Lee

UiTM, Universiti Putra Malaysia

Risk management report (henceforth, RMR) is an audited section in corporate annual reports of companies which offers significant information on the risks they experience and how the risks are mitigated. Today, strict regulation to publish RMRs on a regular basis has been enforced on all public listed companies in the world including banks. Since Malaysia is rather new in the enforcement of RMRs, it would be interesting to see how linguistic features vary in RMRs by Malaysian banks and American banks which have been pioneering the practice. Therefore, this study conducts a comparative multi-dimensional analysis of the linguistic features of RMRs in the corporate annual reports produced by Malaysian and American banks between 2016 and 2020 using Multi-dimensional Analysis Tagger. The results indicate that RMRs by both countries are informational, non-narrative, context independent, less persuasive, abstract and well-planned. Further statistical investigation using SPSS show that RMRs by American banks are more interactional, affective and comprise more elements of persuasion while reports by Malaysian banks are more context independent. Key linguistic features which contribute to these differences will be discussed and elaborated in this study. The findings provide better comprehension of the genre features of RMRs by Malaysian and American banks in their annual reports. The present study also benefits readers, stakeholders and researchers of RMR.



The Design and Development of MUET Corpus

Noorli Khamis

Universiti Teknikal Malaysia Melaka (UTeM)

Malaysian University English Test (MUET) has been used as the benchmark to determine the English language proficiency of students for admission to public universities and colleges in Malaysia. Recent proactive steps, taken by the administrator of MUET, the Malaysian Examination Council (MEC), which include the introduction of the computer-based test version of MUET, i.e. MUET on Demand (MoD) and the alignment of MUET syllabus to the Common European Framework Reference (CEFR) promise a rise in the demand for the test. With the recognition from the Cambridge English Language Assessment (CELA), MUET now sets out to be an internationally-accepted qualification for university entry. As such, looking at the relevance of MUET teaching and learning to cater to the growing demand, both locally and internationally, this paper introduces MUET Corpus as a resource for the language research and learning of MUET. Corpus-based approach has been regarded as a valuable method in extracting useful input for language description in language learning. The corpus is a collaboration project with MEC, and it contains the written texts of the test papers for all the skills – Reading, Listening, Writing, and Speaking. Being an open corpus, currently, MUET Corpus collects the papers from 1999 until 2020. This paper reports the design and development of the corpus. The discussion includes the challenges faced in the development of the corpus, which provides critical considerations on corpus construction issues.



A Corpus Analysis of Metaphorical Expressions in Covid-19 Open Research Dataset (CORD19)

Shazia Riaz Dar

GIFT University

Metaphors play a significant role in shaping our thoughts and processes of communication. They are extensively used in everyday language and hence are an indispensable part of shaping human thought and language structures. Therefore, amid the current situation of Covid-19, this paper investigated variation in language use particularly focusing on metaphors to describe this pandemic. Adhering to a mixed-method, the corpora of Covid-19 Open Research Dataset (CORD-19) were selected since it was openly available in the Sketch Engine database. Quantitatively, the analysis of metaphorical expressions was carried out with the help of corpus tools. Furthermore, the concordances of the selected metaphors were discussed with a critical lens. The findings of the current work showed that people made use of metaphorical expressions from the domains of war, crime, and calamity to highlight the gravity of the pandemic. Moreover, the results also showed that these metaphors denote aspects of real war, crime, and catastrophe and hence demand urgent action in an emergency. They also brought to light that the use of such metaphors inflicted damage on the thinking processes of the people.



A Corpus-Assisted Critical Discourse Analysis of Malaysian Religious COVID-19 Anti-Vaccine Telegram Posts

Mohd Nazriq Noor Ahmad
Universiti Malaya

Religious anti-vaccine movements have existed in Malaysia before the COVID-19 pandemic. These movements continue to disseminate their ideologies despite these activities being declared as illegal by the Malaysian government. They spread vaccine hesitancy through various Telegram channels and websites which they operate. This research aims to provide a better understanding of how Malaysian religious anti-vaccine groups disseminate their ideologies to influence others through posts on Telegram. The research framework utilizes corpus-driven Critical Discourse Analysis, adapting the Corpus-Assisted CDA approach of Baker et al. (2008) along with Reisigl and Wodak's (2016) Discourse-Historical Approach (DHA). The chat history of selected prominent Malaysian religious anti-vaccine Telegram groups are exported and compiled into the form of a corpus. This research utilizes corpus methodologies to examine keyword frequencies, lexical patterns and collocations in the corpus related to the groups' discursive strategies present in the data. It is hoped that this research would provide new ways to utilize corpus linguistics in the analyse data from social networking platforms such as a Telegram."



A Corpus-Driven Analysis of Lexical Variation and Change in Malaysian Twitter.

Noraishah Gulnazir

Universiti Kebangsaan Malaysia

American English is regarded as the hub for global English in Mair's theory of The World System of Englishes. As American English is deemed as a hyper-central language, naturally, the American lexical items are expected to be prevalent in other varieties of English in the world. The advent of social media networks such as Twitter has proliferated the frequency and speed at which lexical items spread among global Internet users. Interestingly, despite adopting British English in the national education system, American English is gaining prominence among Malaysians due to the global spread of American English through the media. In spite of the massive influence of American English, there is a dearth of research on how American English is affecting other varieties of English, especially Malaysian English. There is a need to examine the role of American English in leading global language variation and change. The current research aims at understanding the influence of American English on Malaysian English, particularly on Twitter. To achieve this, the emerging American lexical item 'lit' will be investigated in terms of its trend of frequency and patterns usage in Malaysian Twitter using a corpus-driven analysis. Concordance, collocational and cluster analysis in the WordSmith Tool will be utilized. Results from this study could reveal the extent to which Malaysian English is influenced by American English in terms of lexical units on Twitter, shedding light on the global transformation of the English language.



Deconstructing News Media Stereotypes: An Analysis of Western News Discourse on The Palm Oil Industry in Southeast Asia

**Ravichandran Vengadasamy & Mohamad Iadid Ashrai
Hassannudin**

Universiti Kebangsaan Malaysia / UiTM

This study attempts to deconstruct the various anti-palm oil narratives generated in the western media in the past decade. According to reports in the western news media, oil palm cultivation has resulted in deforestation, loss of habitat for flora and fauna, reduced biodiversity, human rights breaches, and land and labour issues. In addition, palm oil-based products have been subjected to various issues pertaining to quality and health. Therefore, the current study analyses a selected sample of news reports in the western media about the palm oil industry and palm oil products to establish the salient points the media are using to discredit the industry. The research method employs a corpus analysis approach that incorporates both lexical and cognitive analysis of the narratives. The western news articles are collected from 2010 to 2021 to analyse the frequency of corpus used according to the respective years. This study employs a corpus-driven analysis approach using Antconc 3.5.8 (Windows) 2019 and NVIVO 1.0 to analyse the corpus of the western media articles on their narratives of the palm oil industry and products. The frequency of corpus generated in Antconc 3.5.8 (Windows) 2019 and the codes from Nvivo Software are analysed to investigate in detail the news discourse used in the shifting trends and sentiments portrayed through adjectives collocations in the western news media about palm oil from year to year. By doing so, the researchers have deconstructed the prevailing western ideology that permeates much of the narrative in their news media about the oil palm industry and palm oil products. The study will reveal the media discourse as stereotyped narratives that form part of a media campaign aimed at discrediting the palm oil industry focusing on Southeast Asia.



MCRS2022

Epistemic Modal Verbs in George W. Bush's Presidential Speeches: Persuading Americans to go to war in Iraq.

Yasir Ahmed Mohammed Ridha Al-Sabbagh

Universiti Kebangsaan Malaysia

As evidenced by the many studies on the Iraq War, the justification for the war remains controversial up to this day. Consequently, several political and linguistic studies have been conducted to account for the issue of the Iraq War and how Bush persuaded the world to remove the Iraqi government. The data of the political speeches in this paper focuses on the Iraq War, a war that spanned 19 years. Based on a corpus of two speech genres - the State of the Union Speeches and the War Speeches quantitative and qualitative investigations of the epistemic modal verbs in these speeches was conducted using Saeed's framework. The findings demonstrate that George W. Bush made persuasive attempts by relying heavily on 10 functions of epistemic modal verbs (EMV) used recurrently in two genres, namely: Anticipation, Assertion, Assumption, Conviction, Hedging, Incitement, Likelihood/ Possibility, Prediction, Tentative Assertion, and Tricky/Elusiveness. These EMVs were meant to persuade rather than inform the target public on Bush's activities and objectives to resist and eliminate Saddam Hussein's regime. The data also demonstrate that explanations beyond epistemic modal verbs generate public agreement and support, reassure the public, and diminish the opponents' concerns.

Tanggapan Penutur Asing Dewasa terhadap E-buku Interaktif berdasarkan Pendekatan Model Penerimaan Teknologi (TAM)

Marene Mohammad

Universiti Teknologi Malaysia

Model Penerimaan Teknologi (TAM) dapat digunakan untuk mengkaji penggunaan teknologi dalam sesuatu pengajaran dan pembelajaran dengan penelitian secara empirikal. Objektif kajian ini untuk mentafsir tanggapan penutur asing dewasa terhadap e-buku interaktif adalah berguna dan mudah diguna sebagai alat pembelajaran sendiri. 18 orang responden dalam kalangan ekspatriat (pegawai dagang) dengan pelbagai latarbelakang terlibat dalam kajian tinjauan yang berasaskan borang soal selidik. TAM digunakan untuk menjawab persoalan-persoalan kajian mengenai tanggapan responden terhadap e-buku interaktif. Dapatan kajian menunjukkan tanggapan yang positif di kalangan ekspatriat terhadap e-buku interaktif iaitu terbukti berguna sebagai alat bantu pembelajaran sendiri dan mudah digunakan dari aspek reka bentuk.



Preferensi semantik: Analisis berasaskan korpus tentang Perubahan Iklim dalam Korpus Hansard Malaysia (MHC)/ Semantic Preference : A Corpus Based Analysis of Climate Change in the Malaysian Hansard Corpus (MHC)

Muhammad Zakwan Mohd Izam

International Islamic University Malaysia

Climate change is a global phenomenon that is no stranger to the eyes of the world's population to the extent that it even made it as one of the topics discussed by the politicians. In the past, studies on the causes and how climate change occurs were often discussed and debated within the scope of science only. However, the discussion on climate change has shifted from the study of phenomena to the study of social, ethical, cultural, communication and even political forms (Hulme 2009, 2013; see also Giddens 2009; Malone 2009). Al Gore, the former vice president of the United States stated in his article in the New York Times entitled "We Can't Wish Away Climate Change; how world leaders failed to muster anything other than a decision to 'take notice' about the intention to act against climate change. This paper looks into the semantic preference of perubahan iklim / climate change in the Malaysian Hansard Corpus (MHC) from parliament 1 (P1) to parliament 14 (P14). A corpus based analysis is used to find the collocates of the targeted words. Thus, by focusing on the collocates of perubahan iklim / climate change, this corpus-driven account will be able to depict the trend of the parliamentary debates in relation to perubahan iklim / climate change and their collocates. This study is expected to provide a clearer view of how climate change is discussed in the Malaysian parliament.



Peranan Kemanusiaan Digital ke Arah Perkembangan Penyelidikan Proses Penterjemahan Digital (DTPR)

Anis Shahirah Abdul Sukur

Universiti Sains Malaysia

Makalah ini meneliti kajian-kajian signifikan tentang penyelidikan proses penterjemahan (translation process research, TPR) serta perkembangannya akibat proses pendigitalan. Kajian TPR atau kini lebih dikenal dengan nama digital translation process research (DTPR) umumnya berfokus pada proses terjemahan manusia dan melibatkan aspek kognitif. Golongan penterjemah menjadi fokus penelitian kerana kajian Bialystok et al. (2013) memberikan maklumat bahawa individu dwibahasa memiliki kemahiran kognitif yang baik. Secara tidak langsung, kajian terhadap proses terjemahan membolehkan maklumat lanjut berkaitan proses psikologi yang berlaku dalam minda penterjemah akan berjaya dibentuk. Kedua, kajian ini akan menambah dapatan umum tentang pemprosesan bahasa dan strategi kognitif yang diaplikasikan oleh penterjemah. Proses yang berlaku dalam minda penterjemah ini dahulunya diteroka menerusi soal selidik, temu bual, dan protokol penzahiran verbal (think-aloud protocol, TAP). Namun, percambahan kajian Kemanusiaan Digital mengakibatkan proses penterjemahan yang dilakukan oleh manusia ini boleh direkodkan secara nyata dan dianalisis secara sistematik melalui perisian pengvisualan data, analisis data dengan menggunakan komputer dan aplikasi multimedia. Oleh itu, tinjauan awal ini diharapkan dapat memberi perspektif tambahan tentang DTPR dan perkaitannya dengan Kemanusiaan Digital.



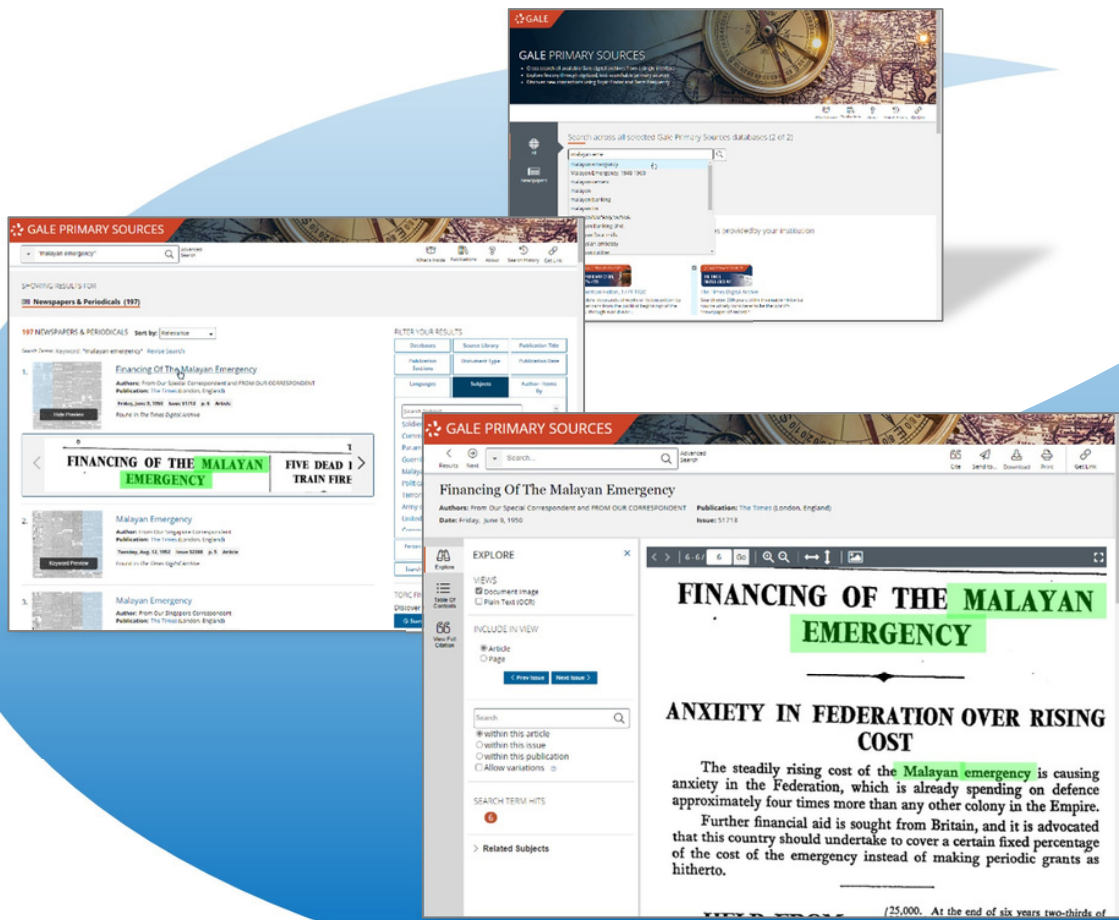
MCRS2022

JAWATANKUASA PENGANJUR SIMPOSIUM PENYELIDIKAN KAJIAN KORPUS 2022

Penasihat	Prof. Dr. Kadaruddin Ayub Prof. Madya Dr Nor Fariza Mohd Nor
Pengerusi Bersama / Co-convenors	Prof. Dato Dr. Imran Ho Abdullah Dr. Azhar Jaluddin
Timbalan Pengerusi I	Dr. Azizul Haji Ismail (DBP)
Timbalan Pengerusi II	Dr. Siti Aeisha Joharry (UiTM/ MCRN)
Setiausaha	Dr. Anis Nadiah Che Abdul Rahman (UKM/ MCRN)
Penolong Setiausaha	Puan Nor Azleenda Abdul Hadi (DBP)
Bendahari	Dr. Intan Safinaz Zainudin
Sekretariat dan Program	Dr. Kesumawati Abu Bakar (Ketua) Prof. Madya Dr. Afendi Hamat Dr Zaharom Mohd Ridzwan Prof Madya Dr Marlyna Maros Pn. Fuzirah Hashim Puan Ruzanna Ruslan (DBP) Puan Aziyanti Abdul Aziz (DBP) Puan Wan Norasikin Wan Ismail (DBP) Puan Nurulhuda Mohd Ali (DBP) Cik Tasha Erina Taufek Encik Yasir Ahmed
Publisiti, Media, Logistik dan Teknikal	Pn. Assmah Osman (DBP) Encik Dinie Asyraf Rodzi (DBP) Encik Mohd Shafik bin Khamshah Encik Mohd Faizal Idzham bin Tajuddin
Sukarelawan	Dr Hazleena Baharun Cik Fareeha Aazam Encik Nazri Bin Mohd Noor Cik Sri Fitriaty binti Mohd Kenali Encik Amir Lukman Abd Rahman

Explore centuries of world history with:

Gale Primary Sources



FREE TRIAL LINK for MCRN: [bit.ly/mcr n-gps](https://bit.ly/mcr-n-gps) Good until 31 Oct 2022

Working with the academic community and archives across the world, Gale digitizes physical materials and apply unique search technology designed to aid discovery in academic research, helping researchers develop new ideas and deliver original research at all levels of study. Start from the source—explore world history with **Gale Primary Sources**.

For more details, contact sean.chia@cengage.com

EMPOWER™ RESEARCH



Gather Analyze

Conduct deep textual analyses with:

Gale Digital Scholar Lab

The image displays three screenshots of the Gale Digital Scholar Lab interface, demonstrating its analytical capabilities:

- Ngrams:** A word cloud visualization for the text "Malayan Emergency 1948-1960". Key terms include "Kuala Lumpur", "abdul rahman", "legislative council", "high commissioner", "federation malaya", "security for", "crime minister", "united states", "india", "british government", "malayan communist", "terrorist", "guerrilla", "guerrillas", "guerrilla warfare", "guerrilla fighters", "guerrilla forces", "guerrilla units", "guerrilla operations", "guerrilla activities", "guerrilla tactics", "guerrilla strategy", "guerrilla warfare", "guerrilla fighters", "guerrilla forces", "guerrilla units", "guerrilla operations", "guerrilla activities", "guerrilla tactics", "guerrilla strategy".
- Sentiment Analysis:** A scatter plot showing the distribution of sentiment scores for the same text. The plot is divided into five regions: Strongly Negative (red), Negative (orange), Neutral (blue), Positive (green), and Strongly Positive (yellow). A tooltip for "Labour Troubles in Malaya" shows a score of -1.667 and a percentage of 17.6%.
- Named Entity Recognition:** A table listing identified entities and their categories. The table includes columns for Entity, Category, and Documents/Count.

Entity	Category	Documents	Count
Britain	Cultural Group	329	602
Federation	Organization	306	574
one	Number	379	567
Malayan	Cultural Group	361	549
first	Position	342	458
Kuala Lumpur	Geo-Political Entity	378	470
Communists	Cultural Group	219	463
yesterday	Date	368	462
London	Geo-Political Entity	275	418
three	Number	298	395
Britain	Geo-Political Entity	171	302
Malay	Cultural Group	185	291
State	Organization	175	269
ahouse	Geo-Political	182	242

FREE TRIAL LINK for MCRN: bit.ly/mcr-n-dsl Good until 31 Oct 2022

By integrating an unmatched depth and breadth of digital cloud-based primary source OCR text with the most popular digital humanities tools, **Gale Digital Scholar Lab** provides a new lens to explore history and empowers researchers to generate world-altering conclusions and outcomes, all without the need of any coding or special skills.

EMPOWER RESEARCH

Organised by



UNIVERSITI
KEBANGSAAN
MALAYSIA
*The National University
of Malaysia*

Fakulti
& Sains Sosial
& Kemanusiaan
Faculty of Social Sciences & Humanities, UKM

pkbl PUSAT KAJIAN
BAHASA &
LINGUISTIK
Centre for Research in Language & Linguistics



Terima kasih *Thank You*



MCRS2022



Bangi-Resort Hotel

Bandar Baru Bangi, Selangor



email

mcrs2022.ukm@gmail.com



Website

<https://mcrs2022.weebly.com>



GALE



MCRN

Malaysian Corpus Research Network