

RESEARCH ARTICLE

Analisis Korpus Jaringan Leksikal Seksual Menggunakan Perisian *Graphcoll* Berdasarkan Laporan Jenayah BH Online (2004–2020)

Muhamad Fadzllah Zaini*¹, Md Zahril Nizam Md Yusoff¹, Darwalis Sazan² dan Farah Fazlinda Mohamad¹

¹Fakulti Bahasa dan Komunikasi, Universiti Pendidikan Sultan Idris, 35900, Tg. Malim, Perak, Malaysia.

²Pusat Pengajian Pendidikan Jarak Jauh, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia.

ABSTRAK – Seksual dalam konteks jenayah merujuk kepada pelanggaran undang-undang seperti yang digariskan dalam Kanun Keseksaan, Akta 792 dan Akta 840, dan sering dikaitkan dengan implikasi serius terhadap kesejahteraan mental dan sosial masyarakat. Kajian ini bertujuan menganalisis pola kolokasi perkataan seksual dalam Korpus Perlaporan Akhbar Jenayah BHonline (2004–2020) bagi memahami gambaran jenayah seksual dalam wacana media. Data korpus terdiri daripada 3,064,134 token, 48,351 jenis kata dan 7,950 teks, dan dianalisis menggunakan pendekatan Statistik Linguistik Korpus melalui perisian #LancsBoxX. Dapatan menunjukkan bahawa perkataan seksual berkolokasi secara signifikan dengan kata seperti kanak-kanak, penderaan, dan gangguan, yang mencerminkan hubungan rapat antara jenayah seksual dan kerentanan mangsa. Kolokasi tersebut dikelaskan kepada beberapa tema utama iaitu konflik [gangguan (11.9), penderaan (11.7), pedofilia (8)], budaya [gejala (8.1)], serta pencirian sifat [kanak-kanak (11.2), pemangsa (8.4)]. Kata lain seperti amang, penganiayaan, bapa, anak, polis, dan guru turut muncul dalam jaringan kolokasi dan menggambarkan konteks pelaku, mangsa serta institusi terlibat. Secara keseluruhan, pola kolokasi ini memperlihatkan bagaimana media membingkai isu jenayah seksual melalui pilihan leksikal yang konsisten. Hasil kajian dapat membantu pihak berkepentingan merangka dasar serta strategi pencegahan yang lebih berkesan dalam menangani jenayah seksual di Malaysia.

ARTICLE HISTORY

Received: 11 October 2024

Revised: 10 November 2025

Accepted: 18 December 2025

Published: 22 December 2025

KATA KUNCI

Jenayah

Linguistik korpus

Seksual

Teks akhbar

Tematik

Corpus-based Analysis of a Sexual Lexical Network Using *Graphcoll* based on BH Online Crime Reports (2004–2020)

ABSTRACT – Sexual acts in the context of crime refer to conduct that violates the law as provided under the Penal Code, Act 792, and Act 840, and are often associated with psychological harm and long-term social implications. This study aims to analyse the collocational patterns of the word sexual in the BHonline Crime News Corpus (2004–2020) to understand how sexual crimes are represented in Malaysian media discourse. The corpus comprises 3,064,134 tokens, 48,351-word types, and 7,950 texts, and the analysis was conducted using a quantitative Corpus Linguistics approach with the #LancsBox X software. The findings reveal that sexual frequently co-occurs with terms such as children, abuse, and harassment, indicating strong lexical associations with vulnerable victims and criminal behaviour. These collocations are classified into three main themes: conflict [harassment (11.9), abuse (11.7), paedophilia (8)], socio-cultural framing [symptom (8.1)], and characterisation [children (11.2), predator (8.4)]. Additional collocates such as molestation, maltreatment, father, child, police, and teacher further illustrate the roles of perpetrators, victims, and institutional actors. Overall, the collocational patterns demonstrate how the media constructs societal perceptions of sexual crime through recurring lexical choices. These insights offer valuable implications for policymakers and relevant agencies in designing more effective strategies for addressing sexual crime in Malaysia.

KEYWORDS

Corpus linguistics

Crime

News text

Sexual acts

Thematic

1.0 PENGENALAN

Media berita memainkan peranan penting dalam pelaporan jenayah dan keadilan jenayah orang ramai, termasuk persepsi tentang bagaimana jenayah biasa berlaku, mengapa jenayah berlaku, dan bagaimana jenayah harus ditangani (Beckett 1997; Haney 2009). Walau bagaimanapun, pelaporan jenayah ini memberi impak kepada pembangunan tamadun dalam kemasyarakatan. Bagi media,

*CORRESPONDING AUTHOR | Muhamad Fadzllah Zaini | ✉ mfadzllah@fbk.upsi.edu.my

peranan dan tanggungjawab memastikan pelaporan berita disampaikan dengan baik dan telus. Setiap hari pelbagai pelaporan jenayah yang disampaikan melalui media berita. Penyampaian yang berkesan dapat memberi impak positif kepada masyarakat. Hal ini dikonsepsikan sebagai pendidikan umum kepada masyarakat mengenai keburukan dan kebaikan dalam soal jenayah yang perlu dihindari.

Menerusi petikan daripada laporan akhbar Berita Harian (11 Januari 2023), Kadar Jenayah Indeks Malaysia menurun 4.1 peratus apabila 50,819 kes direkodkan pada tahun 2022 berbanding dengan 52,974 pada tahun 2021. Perkara ini menunjukkan Polis Diraja Malaysia (PDRM) berjaya mengekalkan senario keselamatan negara pada tahap yang baik. Menerusi kenyataan media Statistik Jenayah, Malaysia 2021 yang dikeluarkan oleh Jabatan Perangkaan Malaysia, keselamatan awam semakin baik dengan penurunan kes jenayah 21.4 peratus namun jenayah siber meningkat 99.5 peratus pada 2020 berbanding 2019. Laporan tersebut, menunjukkan lapan (8) statistik jenayah terpilih iaitu jenayah indeks, dadah, rasuah, alam sekitar, kesalahan trafik, jenayah siber, jenayah maritim dan pemulihan. Kesalahan Trafik dan Jenayah Siber merupakan statistik baharu dalam laporan tersebut. Penyusunan statistik ini adalah berpandukan kepada konsep dan garis panduan daripada International Classification of Crime for Statistical Purpose Version 1.0 yang diterbitkan oleh United Nations Office on Drugs and Crime (UNODC).

Perspektif jenayah dapat dilenturkan dalam kajian bahasa terutama pemakaian reka bentuk linguistik korpus. Linguistik korpus banyak memandu pengkaji dalam mencari perincian serta konteks sebenar yang berlaku menerusi data teks. Daripada data teks ini, pelbagai sumber jenayah dilaporkan terutama teks yang melibatkan berita/akhbar (Internet/cetakan). Kajian Parvareh (2023) menunjukkan beberapa kata yang digunakan dalam ungkapan kebencian seperti, *shit, smuggling drugs/opium, Afghani, blood*, dan beberapa perkataan lain. Kajian ini memperlihatkan konteks penggunaan kata dalam ayat bagi menonjolkan dapatan ungkapan kebencian. Di samping itu, terdapat juga kajian yang melihat kata melalui teks undang-undang jenayah seperti *shall, should, will, would, must, can, could, may, dan might* (Farhan, 2023). Selain itu, terdapat juga kajian yang melihat kata kunci berdasarkan bidang jenayah siber seperti kata, *information, security, cybersecurity, cyberspace, network, data, internet, services, dan economic infrastructure*. Kajian sedemikian membawa perbincangan secara teliti dalam unjuran kata dalam kajian melibatkan jenayah. Secara tidak langsung, kajian ini membantu dan memudahkan saluran maklumat diberikan kepada masyarakat umumnya dan sidang keselamatan dalam mengklasifikasikan sesuatu kata dalam bahasa dan linguistik.

Kajian eksplorasi teks juga telah banyak diintegrasikan bersama linguistik korpus. Lanjutan itu, kajian sedemikian selari dengan penggunaan korpus bersumberkan akhbar atau media. Seperti kajian Liu dan Huang (2023) yang mengkaji kerangka tanggungjawab perubahan iklim dalam akhbar Chinese dan Amerika. Melalui kajian ini eksplorasi teks dengan berpandukan dan berasaskan korpus digunakan untuk melihat trend kekerapan kata yang berlaku sebagai pembuktian. Isu tersebut juga pernah dibahas oleh Collins and Nerlich (2016), Gu, (2023), Kramar (2023), Osama (2023) pelbagai perspektif perubahan iklim diunjurkan oleh teks akhbar, dari kata sehingga kolokasi memperjelaskan tingkah laku yang dilaporkan. Kajian bagi kata *seksual* jarang dibahas berdasarkan data akhbar dan pendekatan korpus. Seperti kajian O'Keeffe dan Breen (2007) telah melihat kes penderaan seksual kanak-kanak berdasarkan pelaporan akhbar. Kajian tersebut memperlihatkan konteks seksual kanak-kanak yang banyak dilakukan oleh ahli keluarga (abang). Selari dengan itu, perkataan *seksual* ini boleh diperhalusi menerusi perspektif akhbar Malaysia (Bahasa Melayu). Signifikan perkataan ini menunjukkan tahap tingkah laku sosial masyarakat yang boleh ditelusuri berdasarkan pelaporan akhbar jenayah.

1.1 *Graphcoll*: Jaringan Visualisasi Kolokasi

Menerusi fenomena semasa, kajian linguistik korpus sentiasa mendapat tempat dalam menunjukkan representatif terhadap sesuatu konteks terutama dalam kajian kata. Oleh hal yang demikian, perbincangan kajian literatur dapat diperhalusi menerusi korpus khusus (data), dan graf Kolokasi serta Jaringan (analisis). Hasil daripada perbincangan kajian literatur ini membentuk persoalan kajian yang membantu fokus makalah ini.

1.1.1 *Korpus Khusus*

Korpus khusus ini digunakan untuk mengkaji aspek kepelbagaian atau genre bahasa tertentu. Sebagai contoh, kajian yang ingin dijalankan berkaitan dengan bahasa daripada akhbar, bahasa yang digunakan dalam karangan akademik, atau dalam perbualan lisan antara lelaki dan perempuan. Jadi pengumpulan teks akan melibatkan ciri yang khusus dan berfokus. Kajian juga boleh meletakkan fitur lain bergantung kepada masa atau tempat (lokasi). Antara korpus khusus yang boleh dirujuk *Corpus British Press 1998-2009* (Baker et al., 2013) yang meneliti lakaran muslim berdasarkan pelaporan akhbar british dalam tempoh tahun 1998 hingga 2009. Selain itu, Michigan Corpus of Academic Spoken English, teks dalam korpus ini mengandungi transkripsi bahasa lisan yang direkodkan di institusi akademik merentas Amerika (Baker, 2006). Di samping itu, pelbagai korpus khusus yang dibangunkan bersumberkan akhbar seperti *Specialized Corpora Digitized Historical Newspaper Archives* (Black, 2023), *Specialized Corpus Spanish and Catalan Newspapers* (Giugliano, 2022), dan *Arabic Newspapers Corpus on COVID19* (El-Kanash & Hamdan, 2023).

Sekilas, penggunaan korpus khusus dalam kajian eksplorasi kata bukan satu perkara baharu. Sebaliknya, sumber data yang khusus sedemikian membantu dalam menyampaikan imej dan mesej kepada kejadian sesuatu isu. Bagi perihal jenayah, kajian seperti Ahmad et al. (2020) telah menunjukkan variasi kata baharu yang boleh dirujuk menerusi seksual terhadap kanak-kanak. Kajian tersebut telah menentang teks undang-undang sebagai data korpus khusus. Bagi kajian ini, menggunakan korpus khusus yang berfokus kepada pelaporan jenayah menerusi akhbar dalam talian. Penetapan masa dibuat berdasarkan data yang besar. Signifikan dalam penggunaan korpus khusus ini, mendahului analisis dengan melihat trend kekerapan kata melalui senarai kekerapan. Senarai kekerapan atau *frequency list* merupakan satu penganalisan yang merujuk kepada kekerapan kata sebagai tingkah laku pelaporan dalam memerihalkan sesuatu perkara. Perwakilan atau representatif bagi sesuatu isu perlu ditetapkan, bagi kajian ini perkataan seksual digunakan untuk melihat lakaran jenayah seksual yang dilaporkan oleh akhbar. Kekerapan ini membantu kebanyakan kajian

linguistik korpus dapat memahami korpus khusus ini. Kajian seperti Yüksel et al. (2022) yang menggunakan senarai kekerapan kata untuk melihat perkataan yang boleh digunakan dalam sesuatu teks. Penyelidikan Szczygłowska (2022) menggunakan senarai kata bagi eksplorasi wacana ahli politik secara khusus. Selari dengan penelitian makalah ini, penggunaan korpus khusus iaitu Korpus Akhbar Pelaporan Jenayah 2004-2020 dapat memperincikan senarai kata yang berlaku dan kolokasi *seksual* yang muncul. Soalan kajian 1 dibentuk bagi melihat apakah senarai kata melalui kolokasi *seksual* yang muncul dalam Korpus Akhbar Pelaporan Jenayah 2004-2020 yang digunakan?

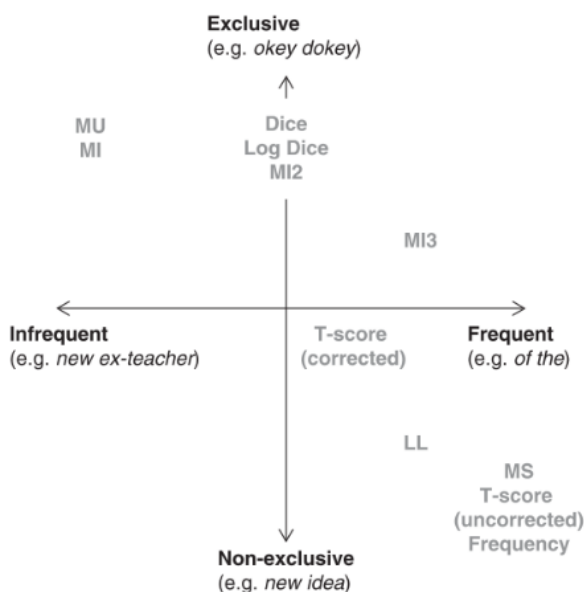
1.1.2 Graf Kolokasi dan Jaringan

Pada dasarnya, kolokasi adalah fenomena yang berkaitan dengan kejadian bersama serta berlaku pengulangan perkataan dalam teks. Penyelidikan kolokasi mewujudkan ruang penelitian kepada asas teks atau ujaran berdasarkan kaca mata daripada hubungan dan kesatuan antara kata (Defanti et al., 2018). Ungkapan yang biasa digunakan oleh linguistik korpus Firth, “*You shall know a word by the company it keeps*” (1957, p.179). Kolokasi ini dianggap sebagai pusat analisis bagi merungkai dan memperincikan kajian kebersamaan kata dalam korpus (Gries, 2013). Menerusi penulisan Brezina et al. (2015) secara tradisinya, tiga kriteria untuk mengenal pasti kolokasi telah dicadangkan. Ini adalah: (i) jarak, (ii) kekerapan, dan (iii) eksklusif. Gries (2013) menunjukkan tiga kriteria lain yang harus dipertimbangkan: (iv) arah, (v) serakan dan (vi) pengedaran token jenis di antara kolokasi. Kebanyakan alat korpus dibangunkan kurang menyediakan platform untuk memperincikan analisis kolokasi seperti graphcoll. (Brezina, 2018) telah memperkenalkan graphcoll menerusi perisian #LancsBoxX untuk meneliti kolokasi. Berdasarkan alat korpus ini, pengujian kolokasi dilakukan berdasarkan variasi ujian untuk menentukan kejadian bersama antara nod dan kata.

Oleh itu, dari sudut analisis, kajian ini menggunakan teknik korpus (Brezina & Gablastova 2018; McEnergy & Hardie 2011) untuk mendedahkan pola kejadian bersama yang kerap dan/atau eksklusif dan membandingkan kekuatan perkaitannya. Seterusnya, untuk memilih ukuran kesatuan, pengkaji perlu menentukan jenis kolokasi yang difokuskan dan diminati (berdasarkan soalan kajian). Kebanyakan ukuran kesatuan ini menonjolkan kolokasi sepanjang dua dimensi utama: kekerapan dan eksklusif. Kekerapan merujuk kepada bilangan kejadian berlaku antara nod dan kata berlaku bersama dalam korpus. Manakala, eksklusiviti merujuk kepada aspek khusus perhubungan kolokasi perkataan berlaku sahaja atau lebih dominan dalam kesatuan atau sama lain (Brezina, 2018). Rajah 1 menunjukkan skala kekerapan dan Eksklusiviti.

Rajah 1

Skala kekerapan dan eksklusiviti



Berdasarkan Rajah 1 ini, antara pengujian yang boleh digunakan untuk mengukur kolokasi ialah MI-skor, logDice dan t-skor (rujuk Jadual 1). Kajian Fontaine, (2017) telah menggunakan MI-skor, logDice dan t-skor untuk melihat kolokasi unik bagi melihat pelaporan mengenai BREXIT dalam akhbar. Keunikan kolokasi memberikan dapatan terhadap perspektif yang khusus bergantung kepada korpus. Kajian Zaini et al. (2020) juga mendapatkan kolokasi ukuran dengan pelbagai dapatan metafora khusus dalam manuskrip Melayu. Oleh itu, persoalan kajian berikut dicadangkan, bagaimanakah kolokasi “*seksual*” yang muncul dalam Korpus Akhbar Pelaporan Jenayah 2004-2020?

Jadual 1

Bentuk fomular bagi tiga (3) pengukuran kolokasi

MI-skor	logDice	t-skor
$MI (Mutual Information) = \log_2 \frac{O_{11}}{E_{11}}$	$logDice = 14 + \log_2 \frac{2 \times O_{11}}{R_1 + C_1}$	$t - skor = \frac{O_{11} - E_{11}}{\sqrt{O_{11}}}$

Hasil daripada pengujian kolokasi ini, nilai tersebut dalam dipaparkan dalam bentuk graf jaringan visualisasi iaitu graphcoll. Graphcoll atau jaringan visualisasi kolokasi ialah teknik tafsiran yang berkuasa sesuai untuk analisis hubungan linguistik yang kompleks dalam data korpus (Defanti et al., 2018). Justeru, persoalan kajian berikut dicadangkan, bagaimanakah bentuk hubungan kolokasi “seksual” yang muncul berdasarkan *graphColl*?

2.0 METODOLOGI

Pangkalan data akhbar dalam talian, Nexis® digunakan untuk mengumpul artikel akhbar yang mengandungi perkataan yang berkaitan dengan Kes, Jenayah dan Polis. Berdasarkan carian ini, kajian ini menggunakan Berita Harian yang merupakan akhbar popular di Malaysia dan berbahasa Melayu di pangkalan data Nexis® antara 1 Januari 2004 hingga 31 Disember 2020 (rujuk Rajah 2). Hasil daripada carian tersebut mengandungi pelbagai artikel akhbar yang memerihalkan soal jenayah yang berlaku di Malaysia. Artikel yang terkumpul sebanyak 7,950 teks dengan melibatkan 3 juta perkataan.

Rajah 2

Carian pangkalan data Nexis®

The screenshot displays the Nexis® advanced search interface. At the top, there is a search bar with the query "kes and jenayah and polis and publication(berita harian)". Below the search bar, there are several filter options on the left side, including "Terms", "Document segments/fields", "Headline", "Title", "Headline and Lead Sections", "Publication", "Date", and "Byline". On the right side, there is a section titled "Use connectors" which lists various search operators: "Exact phrase", "and", "or", "and not", "/n", and "!", each with a brief description of its function. Below this section, there are "Segment Examples" which show sample search results with highlighted terms.

Secara keseluruhan, kajian ini melibatkan metodologi secara gabungan iaitu berpandukan korpus (*corpus driven*) dan pendekatan berasaskan korpus (*corpus-based approaches*). Berdasarkan dua pendekatan tersebut diimplimentasikan dalam Statistik Linguistik Korpus (Brezina, 2018). Hal ini dapat melihat analisis dilakukan secara berpandukan data yang memungkin pelbagai dapatan kolokasi kata, kemudiannya menggunakan hipotesis sedia ada untuk menguji kehadiran kolokasi dalam perspektif jenayah (Baker et al., 2013). Contohnya, menerusi kajian sebelum ini penggunaan kata kunci dan senarai kekerapan kata telah digunakan dalam mencari pelbagai isu menerusi sarjana lain atau kajian rintis yang telah dilakukan (Zaini et al., 2022). Untuk kajian ini fokus diberikan kepada kata seksual, namun perkataan ini bukan berapa pada senarai kekerapan yang tertinggi, perkataan ini mempunyai hubungan kepada percambahan atau perkaitan tindakan atau definisi di bawah jenayah seksual (Pizarro, 2019). Oleh itu, dari segi kefahaman bagaimana set jenayah seksual ini dibina dalam akhbar, perkataan *seksual* memungkinkan kerelevanan dalam membentuk analisis kajian ini.

Langkah pertama, untuk menjawab soalan kajian pertama iaitu Apakah senarai kekerapan yang berkolokasi dengan perkataan seksual dalam Korpus Akhbar Pelaporan Jenayah 2004-2020 yang digunakan? Fokus diberikan kepada hasil kolokasi secara senarai kekerapan kata yang saling berhubung dengan perkataan seksual. Penelitian dari segi kata kandungan secara asas digunakan iaitu kata nama, kata kerja dan kata adjektif (Abrusán et al., 2018). Kata nama seperti penderaan, pesalah, pornografi, pemangsa, kesalahan, yang berkolokasi dengan perkataan seksual. Kata kerja juga dilihat berkolokasi secara bersama dengan perkataan seksual seperti rogol, amang, meliwat, merogol, dan lain-lain kata. Dengan penjanaan daripada alat korpus iaitu perisian #LancsBoxX telah menunjukkan representasi perkataan *seksual* digunakan dalam pelbagai bentuk dan konteks.

Langkah keduanya, menerusi perisian #LancsBoxX pengujian dilakukan berdasarkan Jadual 1. Perkataan seksual dijana secara serentak dengan pengujian MI-skor, logDice dan t-skor bagi melihat representasi yang mungkin berlaku. MI-skor atau Mutual Information, digunakan untuk mengira bilangan atau kejadian sebenar sesuatu perkataan terhadap bilangan kekerapan perkataan itu diramalkan akan berlaku (Smyth, 2016). Bagi logDice pula, untuk menilai kekuatan kolokasi antara dua leksikal saiz korpus tidak menjejaskan pengiraan (Zaini et al., 2020). Selain itu, pengujian t-skor hanya menunjukkan perkataan yang penting bagi perkataan nod, bukan perkataan yang penting bagi perkataan nod (Rosli et al., 2022). Selepas pengujian ini dilakukan representasi kolokasi perkataan seksual akan memperincikan dapatan. Perincian ini dilakukan berdasarkan klasifikasi tema yang boleh dilakukan untuk menentukan kebarangkalian senario yang berlaku dalam jenayah seksual (Baker et al., 2013).

Langkah ketiganya, hasil pengujian ini memberikan graf visualisasi jaringan kolokasi. Graf ini dinamakan sebagai GraphColl. Graf ini menunjukkan plot kolokasi dan perkataan seksual. Plot graf ini disusun berdasarkan nilai signifikan (nilai ambang) bagi setiap pengujian yang digunakan. Selain itu, graf ini juga menggunakan kedudukan tingkap kiri dan kanan bagi melihat konteks penggunaan

perkataan dalam kolokasi. Penandaan melalui plot berwarna juga ditonjolkan dalam graf ini secara automatik bagi membezakan perkataan yang kerap dan eksklusif. Pembuktian konteks dalam pemakaian kata yang berkolokasi dipersembahkan dalam *Keyword In Context* (KWIC) atau Kata Kunci dalam Konteks.

Meskipun kajian ini berasaskan data sekunder yang diperoleh daripada teks akhbar yang telah melalui proses penerbitan formal dan berada dalam domain awam, pertimbangan etika tetap diberikan keutamaan bagi memastikan integriti penyelidikan dipelihara. Analisis linguistik yang dijalankan tidak melibatkan interaksi langsung dengan responden manusia, dan oleh itu tidak mengandungi sebarang pengumpulan data peribadi atau pendedahan maklumat sulit di luar apa yang telah tersedia secara terbuka. Walaupun laporan jenayah seksual lazimnya melibatkan maklumat sensitif, kajian ini tidak mengekstrak, mengulang atau menonjolkan identiti individu, sebaliknya memberi tumpuan eksklusif kepada corak leksikal dan kolokasi sebagai unit analisis. Semua kandungan sensitif dirawat dengan pendekatan berhati-hati (*ethical sensitivity*) bagi mengelakkan interpretasi yang boleh menormalisasi keganasan, menstigmakan mangsa, atau mencetus implikasi etika yang tidak wajar. Kajian ini mematuhi prinsip asas etika penyelidikan termasuk *non-maleficence*, kebertanggungjawaban penyelidik, ketelusan metodologi, dan penghormatan terhadap wacana, selaras dengan kod etika penyelidikan sosial dan standard akademik antarabangsa. Justifikasi penggunaan data akhbar dalam kajian ini terletak pada kepentingannya dalam memahami konstruksi wacana jenayah seksual dalam media secara empirikal, sekali gus menyumbang kepada pengetahuan ilmiah tanpa menimbulkan risiko etika terhadap mana-mana pihak.

3.0 DAPATAN

Hasil daripada analisis kajian, dapatan kajian disusun berdasarkan soalan kajian.

3.1 Kolokasi perkataan seksual

Kolokasi perkataan *seksual* ini telah menggunakan 207 teks daripada 7,950 teks keseluruhan. Kolokasi ini juga melibatkan 661 (215.72) hits, serta merangkumi 1,518 kolokasi meta dan 190 kolokasi yang signifikan dengan tiga pengujian (rujuk Jadual 1). Daripada hasil janaan perisian #LancsBoxX telah mempamerkan beberapa dapatan statistik iaitu kolokasi, serakan, kekerapan (kolokasi)/*freq.(collocation)*, kekerapan (korpus)/ *freq.(corpus)*, logDice, MI-skor, dan T-skor. Jadual 2 menunjukkan dapatan bagi kolokasi perkataan *seksual*.

Jadual 2

Kata berkolokasi perkataan seksual yang signifikan

Kolokasi	Serakan	Kekerapan (Kolokasi)	Kekerapan (Korpus)	log Dice	MI	T-score
gangguan	-0.86	102	229	11.9	11	10.1
penderaan	-0.65	102	356	11.7	10.4	10.1
kanak-kanak	0.73	232	2,646	11.2	8.7	15.2
fizikal	-0.13	48	300	10.7	9.5	6.9
serangan	-0.85	27	356	9.8	8.5	5.2
pesalah	-0.75	24	437	9.5	8	4.9
eksploitasi	-0.73	15	55	9.4	10.3	3.9
seksual	0	22	661	9.1	7.3	4.7
keganasan	-0.14	21	765	8.9	7	4.5
pornografi	-0.6	10	61	8.8	9.6	3.2
jenayah	-0.88	226	18,542	8.6	5.8	14.8
nurin	0.64	11	292	8.6	7.4	3.3
pemangsa	-0.71	7	12	8.4	11.4	2.6
siber	0.27	11	456	8.3	6.8	3.3
kesalahan	-0.59	39	3,396	8.3	5.7	6.1
kesedaran	0	10	418	8.2	6.8	3.1
dewasa	-0.14	7	129	8.2	8	2.6
penganiayaan	-0.67	6	43	8.1	9.3	2.4
gejala	-0.4	10	552	8.1	6.4	3.1
pedofilia	0.67	6	105	8	8	2.4
grooming	0.6	5	7	7.9	11.7	2.2

Berdasarkan Jadual 2, menunjukkan kata yang digunakan secara kerap bersama kolokasi perkataan *seksual*. Kata yang paling signifikan mengikut ukuran MI-skor, logDice dan t-skor ialah kata *gangguan*. Walaupun unjuran dapatan kata ini agak kecil, namun data set kolokasi ini memberi gambaran kepada perspektif *seksual* yang berlaku secara asas.

Walaupun bagaimanapun, kemungkinan jika kolokasi yang kurang kerap juga dipertimbangkan, gambaran yang berbeza mungkin didedahkan (Baker, 2004). Jadi, kesemua 1,518 kolokasi telah diperiksa dan dimasukkan ke dalam kategori tematik melalui analisis konkordans. Ini mendedahkan bahawa kolokasi kata *seksual* tergolong dalam sebilangan kecil tema (rujuk Jadual 3). Perlu dijelaskan bahawa tema tidak begitu bergantung pada makna kamus kata nama yang digabungkan, sepertimana pada topik yang diindeks dalam data korpus. Sebagai contoh, perkataan *kedai* mungkin tidak serta-merta kelihatan sesuai dengan tema 'sifat mencirikan/membezakan', tetapi apabila sesuatu kata mempertimbangkan cara kata kerap digunakan dalam konteks '*pertudahan/dakwaan seksual*', ia nampaknya mempunyai fungsi pembezaan.

Apa yang tercetus daripada penilaian tema dalam bahagian ini ialah penyampaian jenayah seksual sebagai populasi homogen yang terlibat dalam konflik, sama ada sebagai isu dewasa atau melibatkan kanak-kanak paling kerap menjadi yang pertama. Seterusnya, isu konflik dikemukakan (secara tersurat atau/dan tersirat) berpunca daripada jenayah seksual. Kategori dibincangkan dengan lebih terperinci di bawah;

Jadual 3

Tema kata yang menyifatkan seksual

Tema dan sub-tema	Contoh kolokasi kata
Konflik	gangguan, penderaan, pornografi, penganiayaan, pedofilia, ancaman, pembunuhan, rogol, seks, siber, amang
Budaya	
i. Pratikal Sosial	rang, undang-undang, akta,
ii. Pendidikan	sekolah, seminar, pelajar,
iii. Pandangan/Atitud/Emosi	siasatan, dakwaan, statistik, emosi, gejala,
Etnik/Entiti Nasional	
i. Populasi	rakyat, masyarakat, orang,
ii. Kawasan/Negara	putrajaya, mahkamah, bukit aman,
iii. Governan	menteri, parlimen, hakim, pdrm, pengarah, jabatan,
Mencirikan / membezakan sifat	
i. Umur/Jantina	kanak-kanak, remaja, wanita, perempuan, dewasa, bayi,
ii. Keluarga/hubungan	bapa, anak, keluarga, isteri
iii. Pekerjaan/peranan	mangsa, pesalah, penjenayah, guru, pemangsa, pesalah, polis
Kumpulan/Organisasi	unit

Berdasarkan daripada Jadual 3, analisis tematik berdasarkan kolokasi perkataan seksual ini telah menunjukkan representatif sifat *seksual* berdasarkan kepelbagaian kata. Analisis ini telah memetakan kolokasi kepada tema dan sub-tema (rujuk Baker et al., 2013).

3.1.1 Tema Konflik

Kolokasi yang berkaitan dengan perkataan *seksual* ini, kebanyakan mempunyai hubungan yang kuat dalam tema konflik. Kata yang berkaitan ini menunjukkan konflik yang berlaku dalam jenayah *seksual*. Perkataan *gangguan*, *penderaan*, *pornografi*, *penganiayaan*, *pedofilia* antara perkataan yang kerap digunakan bersama perkataan *seksual* (rujuk Rajah 3).

seperti komputer riba, tablet dan telefon pintar di negara ini memudahkan kanak-kanak terdedah kepada jenayah **siber**, termasuk **eksploitasi***seksual*****. Biasanya, orang dewasa akan berkawan dan memenangi kepercayaan mereka dengan memberi perhatian seperti bersimpati dengan masalah atau memberi pelbagai pujian, hadiah, wang atau janji-janji manis. Akhir sekali, pedofilia akan cuba eksploitasi persahabatan itu secara seksual dan berterusan hingga membabitkan pornografi atau perdagangan seks. (Berita Harian, 20 Mei 2016)

Jenayah dengan pengaplikasian teknologi dapat dilihat jejak perkembangan dan mendadak memberi impak terutama kepada masyarakat. Kemunculan perkataan *siber* berkolokasi dengan *seksual* memberi gambaran kepada ketenatan manusia dalam memanfaatkan teknologi melalui peranti untuk memuaskan nafsu mereka. Perubahan landskap ini memberi impak jangka masa panjang dalam jenayah *seksual* bermula dengan eksploitasi *seksual* sehingga membentuk *pedofilia*, *pornografi* dan *perdagangan seks*.

3.1.2 Tema Budaya

Tema ini terdiri daripada kata yang berkaitan dengan budaya yang dilihat sebagai satu set kekaburan sikap, kepercayaan, konvensyen tingkah laku, dan andaian dan nilai asas yang dikongsi oleh sekumpulan orang, dan yang mempengaruhi tingkah laku setiap ahli dan tafsiran setiap ahli tentang 'makna', dan 'tingkah laku orang lain' (Spencer-Oatey 2000: 4). Bagi sub tema pratikal sosial, kolokasi yang memberi gambaran kepada aspek ini seperti *rang*, *undang-undang*, *akta*. Kepercayaan perundangan melindungi budaya dan sosial memperlihatkan penguatkuasaan undang-undang terhadap jenayah *seksual*. Manakala, sub tema pendidikan memperlihatkan peranan *sekolah*, *seminar*, *pelajar* diketengahkan. Kebersamaan perkataan *seksual* ini memperlihatkan dua pembahagian negatif dan positif. Sekolah merujuk kepada institusi pendidikan peringkat rendah, menengah dan prauniversiti. Namun, makna dan tingkah laku negatif juga menjejaskan pendidikan (rujuk DATA KWIC di bawah). Usaha pihak kerajaan terutama badan penguatkuasa undang-undang dalam menangani jenayah *seksual* ini juga dilaporkan melalui akhbar seperti *Seminar Jenayah Seksual Kanak-Kanak: Hentikan!!*. Hal ini kerana, kes sedemikian memberi tamparan hebat kepada masyarakat dalam membangunkan generasi akan datang.

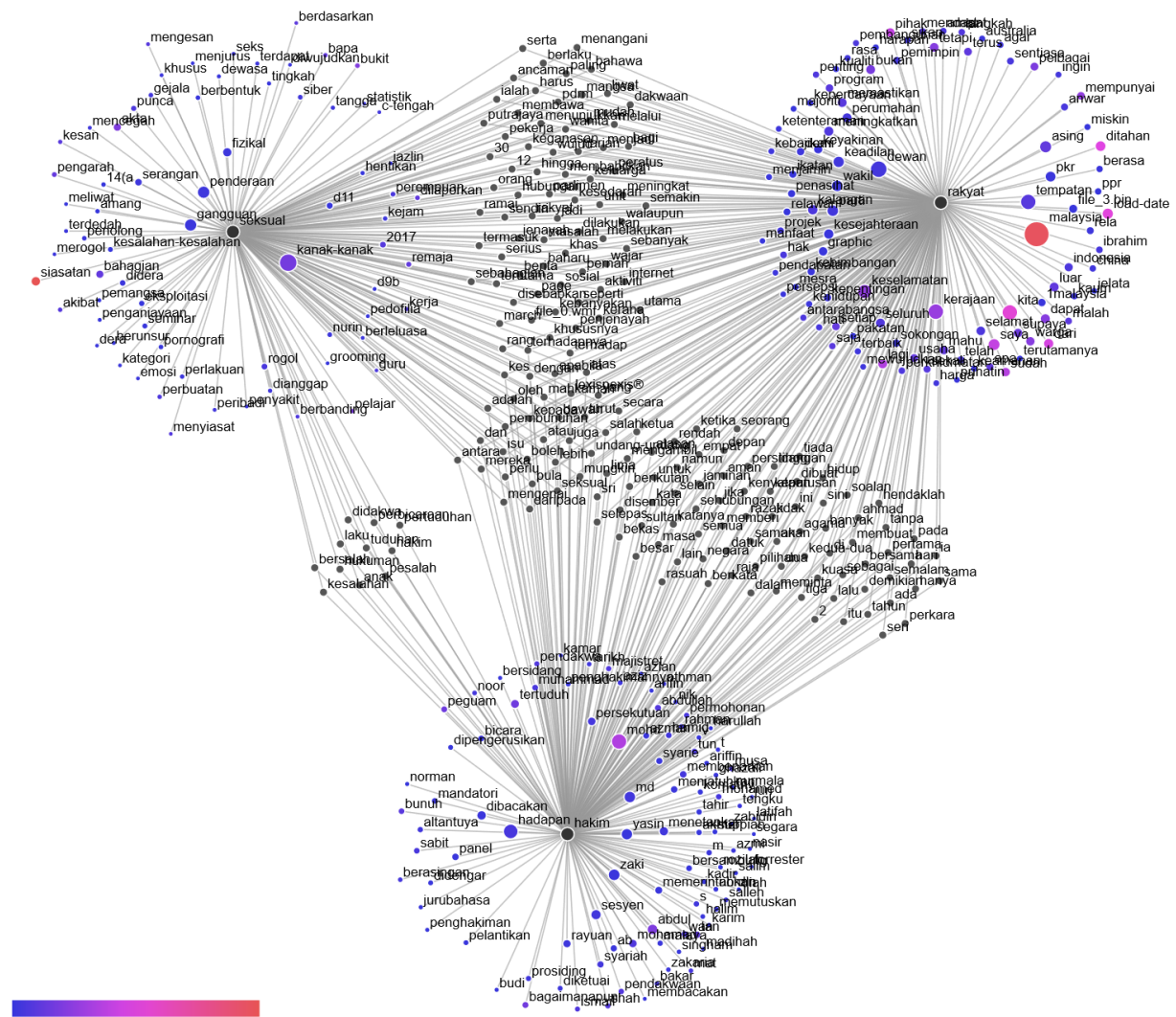
DATA KWIC 4 : punca utama mereka terbabit dalam jenayah itu selepas terpengaruh dengan media sosial seperti *Facebook* dan laman sesawang yang memaparkan video dan bahan lucah di internet. Agensi berita di London melaporkan kajian terbaharu *The Journal of Pediatrics* mendapati perbuatan mengirim pesanan serta gambar lucah secara elektronik atau *sexting* berkemungkinan mendorong aktiviti **seksual** membabitkan bukan saja **pelajar sekolah** tinggi, malah **pelajar sekolah** rendah berusia 12 hingga 14 tahun. Jurnal berkenaan mendedahkan kajian yang dijalankan ke atas 420 responden gred tujuh di lima sekolah rendah awam di bandar Rhode Island, Amerika Syarikat mendapati pelajar cenderung untuk bercium, melakukan seks oral atau hubungan seks berbanding. (Berita Harian, 15 Januari 2014)

DATA KWIC 5 : cuma akan ditambah baik dan diperincikan. Misalnya dalam Akta Kanak-Kanak 2001, sudah ada elemen untuk kesalahan pornografi dan persiapan awal pelaku. Undang-undang lebih lengkap "Namun bagi undang-undang baharu, ia akan diperincikan lagi dengan lebih jelas dan lengkap supaya dapat mengheret pesalah ke muka pengadilan," katanya pada sesi perbincangan **Seminar Jenayah Seksual Kanak-Kanak: Hentikan!!**, semalam. Sementara itu, Pegawai Bahagian Siasatan Seksual, Wanita dan Kanak-kanak Bukit Aman, Deputy Superintendan Saroja Egamparan, mengakui polis menghadapi cabaran dalam siasatan kes jenayah seksual ke atas kanak-kanak orang kurang upaya (OKU). "Ini kerana kekurangan mereka untuk bertutur. Biasanya kami akan mendapatkan keterangan menerusi rakaman video dan memanggil. (Berita Harian, 15 Mac 2017)

Berdasarkan budaya juga, sub tema pandangan/attitud/emosi memberi memberi penekanan melalui kata *siasatan*, *dakwaan*, *statistik*, *emosi*, *gejala*. Perkataan sedemikian, merupakan satu perkataan yang sering digunakan dalam wacana jenayah terutama jenayah seksual. Rangkuman kolokasi sedemikian memerihalkan hasil atau prosedur tindakan. Rajah 4 menunjukkan jaringan visualisasi kolokasi bagi kata *seksual*, *pelajar*, *seminar*. Kolokasi ini memperlihatkan tindakan atau bantuan daripada aspek pendidikan secara langsung. Pelbagai kempen, seminar dan aktiviti berpendidikan dilakukan untuk menangani kes jenayah seksual berkembang.

Rajah 5

Jaringan visualisasi kolokasi seksual-hakim-rakyat



3.1.4 Mencirikan/Membezakan sifat

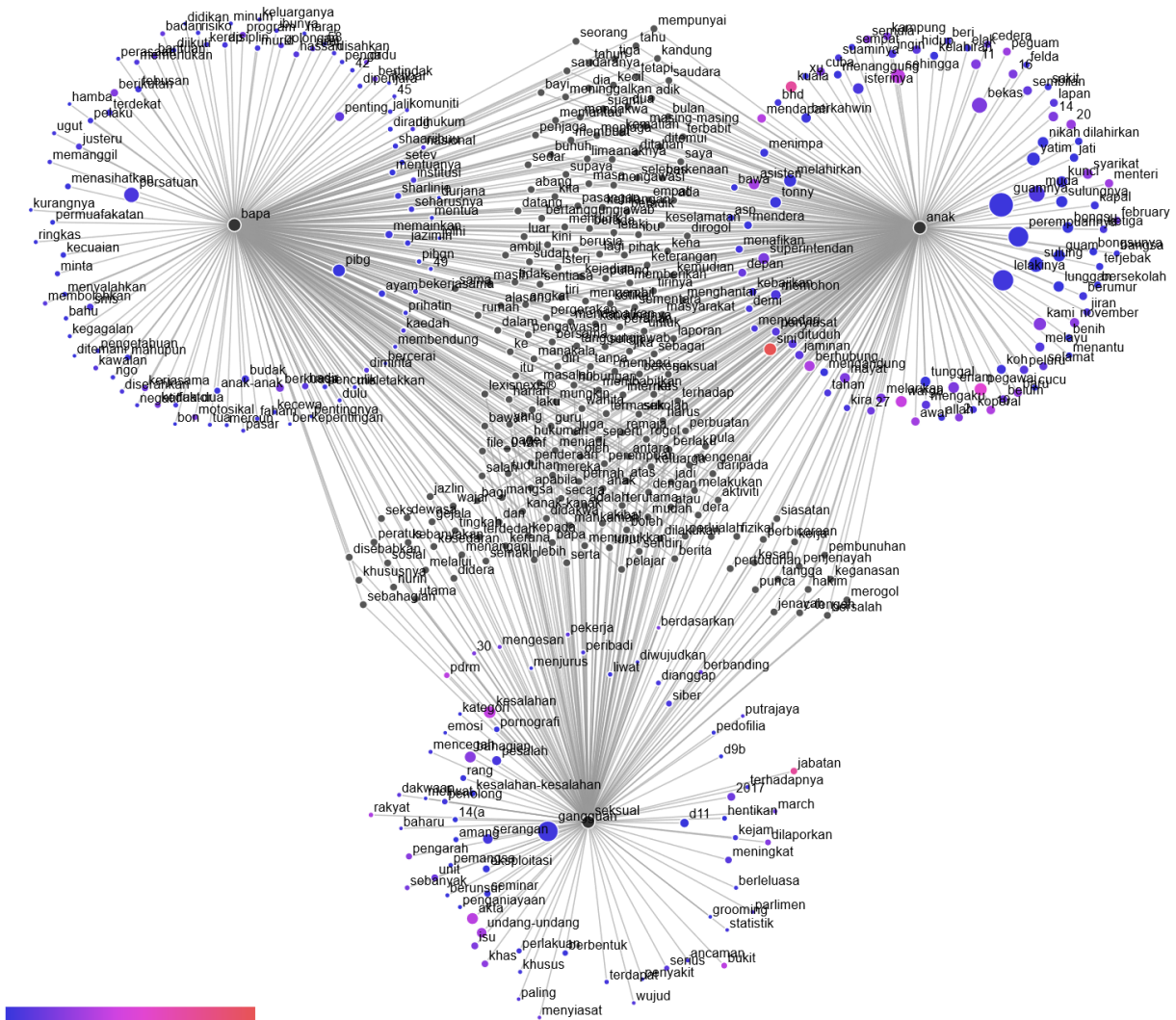
Tema ini merupakan dominan kepada penerangan inti bagi jenayah seksual. Penglibatan sub tema dari segi umur/jantina, keluarga/hubungan, pekerjaan/peranan. Pertama, sub tema umur/jantina yang memerihalkan penglibatan kata yang berkolokasi dengan perkataan seksual. Gambaran ini menjelaskan trend semasa yang melibatkan umur/jantina. Kata yang berkolokasi kanak-kanak, remaja, wanita, perempuan, dewasa, bayi. Kanak-kanak, remaja dan bayi menjadi target utama dalam jenayah seksual. Kata yang menunjukkan jantina seperti wanita, perempuan, lelaki. Dominasi pemangsa atau mangsa atau penjenayah tidak dapat dibandingkan, kerana penglibatan dua entiti (perempuan/lelaki) dalam jenayah seksual sentiasa menjadi fokus. Tetapi, kata wanita, perempuan lebih terleser sebagai mangsa dan lelaki sebagai pemangsa. Apabila mengulas soal mangsa, pemangsa ini banyak kata lain juga yang terlibat seperti pesalah, penjenayah yang mempunyai peranan dalam ruang jenayah sosial. Peranan ini diperihalkan sebagai satu sub tema di bawah pekerjaan/peranan. Peranan merujuk kepada tugas dalam lingkungan jenayah seksual. Sub tema pekerjaan yang berkolokasi dengan kata seksual seperti guru dan polis. Polis mempunyai “cop culture” atau “budaya kepolisan” yang mempunyai pelbagai sisi sama ada negatif atau positif yang bergantung kepada kitaran pekerjaan tersebut.

DATA KWIC 6 : Dua anggota polis ditahan selepas dipercayai melakukan ***seks*** oral, mencabul dan memaksa pasangan kekasih melakukan hubungan ***seks*** di Pengkalan Chepa di sini. Selain merakam kejadian itu, dua anggota berusia 31 dan 35 tahun itu juga dipercayai memeras ugut sebanyak RM1,250, selain menumbuk seorang daripada mangsa. Ketua Polis Kelantan, Deputy Komisioner Shafien Mamat, berkata kejadian berlaku jam 10.45 malam Jumaat lalu ketika dua anggota polis berkenaan sedang melakukan rondaan, sebelum terserempak dengan pasangan kekasih berusia 20-an yang berada dalam sebuah kereta di tepi jalan. (Berita Harian, 9 Disember, 2020)

DATA KWIC 7: mempunyai sebarang interaksi dengan kanak-kanak. Amalan memindahkan guru ke sekolah lain di mana suspek masih berinteraksi dengan kanak-kanak harus dihentikan. Lindungi hak kanak-kanak Apabila mengimbangi keselamatan kanak-kanak dengan hak suspek, hak kanak-kanak harus diutamakan. Hak tertuduh masih dilindungi sekiranya

Rajah 7

Jaringan visualisasi kolokasi seksual-bapa-anak



Rajah 7 memperincikan jaringan kolokasi perkataan seksual bersama kata *bapa*, *anak*. Melalui jaringan ini banyak kata yang membawa konsep negatif terhadap dua entiti kata tersebut. Seperti kata *rogol*, *ancaman*, *seks*, *merogol*, *pembunuhan*, *didera*, *keganasan*, *eksploitasi*, *serangan*, *pesalah*, *penderaan*, *gangguan*. Sepatutnya, peranan bapa sebagai ketua keluarga yang melindungi ahli serta bertanggungjawab menjaga keselamatan. Malah sebaliknya belaku dengan kehadiran set kata negatif ini memberi imej yang buruk terhadap bapa (ketua keluarga).

3.1.5 Kumpulan/organisasi

Tema ini serupa dengan yang sebelumnya, kerana kata *seksual* mencirikan entiti, walaupun secara kolektif. Walau bagaimanapun, penggabungan itu layak dalam kategori yang berasingan kerana kumpulan/organisasi mengindeks matlamat kolektif dan usaha aktif kumpulan/organisasi, bukannya atribut tipikal. Sehubungan itu, tema ini disambungkan kepada 'entiti etnik/kebangsaan'. Apabila rujukan adalah kepada organisasi sebenar, kolokasi itu sendiri adalah 'neutral' (iaitu ia hanya mengindeks organisasi tertentu). Kata yang berkaitan ialah *unit*. Berikut merupakan bentuk pemakaian kata *unit* dalam kolokasi perkataan *seksual* (Rujuk Rajah 8);

DATA KWIC 9 : termasuklah mengadakan kempen bersepadu yang hebat untuk meningkatkan kesedaran dan rasa hormat terhadap hak wanita dalam masyarakat; program latihan PDRM perlu memberi lebih penekanan pada hak wanita dan layanan terhadap suspek dan mangsa wanita; menubuhkan Bahagian Siasatan Jenayah Wanita dan Kanak-kanak yang lebih besar dan lebih lengkap daripada **Unit Siasatan ***Seksual***** yang ada sekarang di Bukit Aman; mematuhi dengan ketat peruntukan undang-undang dan PTKPN berkaitan dengan pengadu wanita dan siasatan terhadap suspek wanita; melengkapkan semua balai polis dengan kemudahan yang mencukupi untuk menangani kes yang membabitkan wanita; menggunakan pendekatan pencegahan yang lebih proaktif terhadap masalah keganasan rumah tangga melalui kepolisan masyarakat. (Berita Harian, 17 Mei 2005)

rendah, sebaliknya perkataan ini kerap digunakan dan semakin bertambah kepenggunaan kata. Oleh hal yang demikian, dengan adanya penyenaaran kekerapan kolokasi ini memberi gambaran kepada konteks jenayah seksual yang berlaku sama ada bercirikan siber, kanak-kanak, liwat atau kegunaan rumah tangga.

Keduanya, bacaan akhbar secara manual dan bacaan mesin dengan menggunakan perisian korpus merupakan dua aspek yang berbeza. Dengan adanya mesin keterbacaan automatik ini akan memberikan set kata dengan nilai kekerapan yang signifikan. Hasil dapatan menunjukkan, kolokasi perkataan seksual ini berlaku dengan signifikan melalui pengujian logDice. Kata yang muncul ialah, *gangguan* [11.9], *penderaan* [11.7], *kanak-kanak* [11.2], *fizikal* [10.7], *serangan* [9.8], *pesalah* [9.5], *eksploitasi* [9.4], *didera* [9.2], *kesalahan-kesalahan* [9.2], *keganasan* [8.9], *rang* [8.8], *pornografi* [8.8], *amang* [8.8], *laku* [8.8], *rogol* [8.7]. Bagi perkataan yang muncul selepas pengujian MI-skor ialah, kata *amang* [12.2], *kesalahan-kesalahan* [12.0], *grooming* [11.7], *gangguan* [11.0], *penderaan* [10.4], *eksploitasi* [10.3], *hentikan* [10.3], *pornografi* [9.6], *fizikal* [9.5]. Sementara itu, bagi pengujian t-skor, beberapa perkataan dikenal pasti berkolokasi secara signifikan dengan perkataan seksual. Antaranya, kata *kanak-kanak* [15.2], *jenayah* [14.8], *kes* [10.3], *gangguan* [10.1], *penderaan* [10.1], *mangsa* [8.6]. Daripada nilai signifikan ini, set kata muncul dengan pelbagai kemungkinan. Setiap pengujian memberikan kata-kata yang berbeza berdasarkan faktor tettingkap kiri-kanan penggunaan dan kekerapan digunakan kiri-kanan. Perkataan yang berkolokasi analisis tematik digunakan untuk klasifikasi tema dalam eksplorasi perkataan seksual. Antara tema yang terlibat ialah konflik (*gangguan, penderaan, pornografi, penganiayaan, rogol, siber*), Budaya (*undang-undang, akta, seminar, pelajar, siasatan, dakwaan*), Etnik/Entiti Nasional (*rakyat, masyarakat, bukit aman, hakim, mahkamah*). Mencirikan/Membezakan Sifat (*kanak-kanak, bapa, anak, mangsa, pesalah, polis, guru*), dan Kumpulan/Organisasi (*unit*).

Ketiganya, jaringan visualisasi kolokasi merupakan satu inovasi yang diperkenalkan oleh Brezina (2018). Jaringan visualisasi kolokasi ini mengambil kira beberapa faktor seperti kedudukan kata (kiri/kanan), nilai signifikan, bilangan tettingkap yang dihadkan (kiri/kanan), bilangan kolokasi, bilangan kekerapan, bilangan keseluruhan korpus (token). Oleh hal yang demikian, jaringan visualisasi ini mempamerkan plot kolokasi kata berdasarkan pengujian yang digunakan (MI-skor, t-skor, logDice). Pameran plot kolokasi ini menunjukkan kehadiran kata yang berlaku dalam lingkungan perkataan seksual. Seperti data kata *seksual-gangguan-penderaan, seksual-pelajar-seminar, seksual-hakim-rakyat, seksual-polis-guru, seksual-bapa-anak, seksual-unit*. Kolokasi ini dapat memberi gambaran kepada konteks jenayah seksual yang dilaporkan dalam akhbar. Peranan setiap anggota masyarakat sama ada badan berkuasa dan ahli keluarga perlu menjaga tertib dan bertanggungjawab. Akhbar merupakan sumber pendidikan umum melalui bahan bacaan yang dekat dengan masyarakat menjadi agen bagi pemupukan kesedaran terhadap jenayah seksual ini.

5.0 KESIMPULAN

Secara kesimpulannya, hasil dapatan ini telah memberi kefahaman kepada kandungan atau trend jenayah seksual dalam pelaporan akhbar. Tema dan sub tema menjadi landasan kepada perbincangan isu jenayah seksual ini. Hubungan antara kata melalui kolokasi menjadi bukti dalam memahami bentuk fenomena yang berlaku seperti konflik, budaya, etnik/entity nasional, mencirikan/membezakan sifat, dan kumpulan organisasi. Peranan senarai kekerapan kata dapat memberi ekspresi kata seksual yang berlaku dan disenaraikan secara terperinci melalui kepelbagaikan kata. Kemudian, hubungan atau jaringan kolokasi kata dapat mengesan konteks dan mewujudkan kefahaman jenayah seksual yang telah berlaku antara tahun 2004 hingga 2020. Akhir sekali, pembuktian hubungan dapat dipersembahkan melalui *GraphColl* yang dijana menerusi perisian LancsBoxX. Pengkelompokan kata berdasarkan tema dapat perincian konteks jenayah seksual yang berlaku dalam pembuktian antara kata. Implikasi daripada kajian ini, menunjukkan fungsi bahasa dan linguistik dalam memberi kefahaman secara rinci bagi jenayah seksual dapat dibuktikan. Bantuan bukti perkataan dapat mendedahkan konteks sebenar yang berlaku tanpa bias dalam pemilihan perkataan. Keterbacaan Mesin Automatik seperti perisian LancsBoxX membantu persembahan kata berdasarkan pengujian yang terpilih tanpa bias manusia. Makalah ini juga, dapat memperlihatkan fenomena yang berlaku bagi maklum balas polisi Matlamat Pembangunan Mampan 16 Keamanan, Keadilan dan Institut yang kuat. Dalam MPM 16 ini penekanan kepada jenayah seksual jelas bagi pembanterasan dan pemeliharaan undang-undang dalam ketamaduan kemanusiaan sejagat.

PENGHARGAAN

Kajian ini telah dijalankan di bawah Skim Geran Penyelidikan Fundamental (FRGS/1/2022/SSI09/UPSI/03/1) yang disediakan oleh Kementerian Pendidikan Tinggi dan diketuai oleh Dr. Muhamad Fadzllah bin Hj Zaini.

KONFLIK KEPENTINGAN

Penulis mengisytiharkan tiada konflik kepentingan.

SUMBANGAN PENULIS

Muhamad Fadzllah Zaini (Penulisan – deraf asal, Penulisan – semakan & penyuntingan, Metodologi)

Md Zahril Nizam Md Yusoff (Analisis formal)

Darwalis Sazan (Konseptualisasi, Pengurusan data)

Farah Fazlinda Mohamad (Pengesahan, Sumber)

RUJUKAN

- Abrusán, M., Asher, N., & van de Cruys, T. (2018). Content vs. function words: The view from distributional semantics. *Proceedings of Sinn Und Bedeutung* 22, 1(269427), 1–21.
- Ahmad, A., Mohamad Mangsor, M., Ardi, N., & Ab. Wahab, A. (2020). Peristilahan Bahasa Melayu dalam Akta Kesalahan Seksual terhadap Kanak-Kanak 2017 (Akta 792). *Jurnal Bahasa*, 20(1), 151–172. [https://doi.org/10.37052/jb.20\(1\)no7](https://doi.org/10.37052/jb.20(1)no7)
- Baker, P. (2004). 'Querying keywords: questions of difference, frequency and sense in keywords analysis.' *Journal of English Linguistics* 32,4, 346–59. <https://doi.org/10.1177/0075424204269894>
- Baker, P. (2006). Using corpora in discourse analysis. In *Continuum* (pp. 1–206). Continuum.
- Baker, P., Gabrielatos, C., & McEnery, T. (2013). Sketching muslims: A corpus driven analysis of representations around the word "Muslim" in the British press 1998-2009. *Applied Linguistics*, 34(3), 255–278. <https://doi.org/10.1093/applin/ams048>
- Beckett, K. (1997). *Making crime pay*. New York: Oxford University Press
- Black, J. W. (2023). Creating specialized corpora from digitized historical newspaper archives: An iterative bootstrapping approach. *Digital Scholarship in the Humanities*, 38(2), 779-797. <https://doi.org/10.1093/lc/fqac079>
- Brezina, V., & Gablasova, D. (2018). The corpus method. In J. Culpeper, J., Kerswill, P., Wodak, R., McEnery, A., & Katamba, F. (2018). *English Language: Description, variation and context (2nd ed.)*. Macmillan Education UK. <https://doi.org/10.1057/978-1-137-57185-4>
- Brezina, V. (2018). Statistics in Corpus Linguistics. In *Cambridge University Press* (First). Cambridge University Press. <https://doi.org/10.1017/9781316410899>
- Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139–173. <https://doi.org/10.1075/ijcl.20.2.01bre>
- Collins, L. C., & Nerlich, B. (2016). Uncertainty discourses in the context of climate change: A corpus-assisted analysis of UK national newspaper articles. *Communications*, 41(3), 291–313. <https://doi.org/10.1515/commun-2016-0009>
- Defanti, T., Grafton, A., Levy, T. E., Manovich, L., & Rockwood, A. (2018). *Lexical Collocation Analysis; Advances and Application* (First). Springer Nature Switzerland AG.
- El-Kanash, H. H., & Hamdan, J. M. (2023). COVID-19 Conceptual Metaphors in Jordanian Political Discourse: Evidence from a Newspaper-based Corpus. *GEMA Online Journal of Language Studies*, 23(1), 93–113. <https://doi.org/10.17576/gema-2023-2301-06>
- Firth, J. R. (1957). *Applications of General Linguistics*. *Transactions of the Philological Society*, 56(1), 1–14. <https://doi.org/10.1111/j.1467-968X.1957.tb00568.x>
- Farhan, A. K. (2023). Divergence in the translation of criminal law: A corpus-based study of prohibition in Iraqi penal code and its English translation. *Ampersand*, 10(December 2022), 100104. <https://doi.org/10.1016/j.amper.2022.100104>
- Fontaine, L. (2017). The early semantics of the neologism BREXIT: a lexicogrammatical approach. *Functional Linguistics*, 4(1). <https://doi.org/10.1186/s40554-017-0040-x>
- Giugliano, M. (2022). Discourses about independence: A corpus-based analysis of discourse prosodies in spanish and catalan newspapers. *Discourse and Communication*, 16(5), 525-550. <https://doi.org/10.1177/17504813221099194>
- Gries, S. T. (2013). 50-Something Years of Work on Collocations. *International Journal of Corpus Linguistics*, 18(1), 137–166. <https://doi.org/10.1075/ijcl.18.1.09gri>
- Gu, C. (2023). 'Climate change concerns human survival. . .and justice in our international community': A corpus-based positive discourse analysis (PDA) of the largest developing nation's global involve/ engagement discourses (re)told in interpreting. *PLoS ONE*, 18(4 April), 1–20. <https://doi.org/10.1371/journal.pone.0277705>
- Haney C. (2009). The social psychology of isolation: why solitary confinement is psychologically harmful. *PrisonServ J*. 181:12–20
- Kramar, N. (2023). Construction of Agency within Climate Change Framing in Media Discourse: a Corpus-Based Study. *Respectus Philologicus*, 43(48), 36–48. <https://doi.org/10.15388/RESPECTUS.2023.43.48.106>
- Liu, M., & Huang, J. (2023). Framing responsibilities for climate change in chinese and american newspapers: A corpus-assisted discourse study. *Journalism*, <https://doi.org/10.1177/14648849231187453>
- McEnery, T and Hardie, A (2012) *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
- O'Keeffe, A., & Breen, M. J. (2007). At the hands of the brothers: A corpus-based lexico-grammatical analysis of stance in newspaper reporting of child sexual abuse cases. *The language of sexual crime*, 217-236. <https://doi:10.1057/9780230592780>
- Osama Ghoraba, M. (2023). Influential Spanish Politicians' Discourse of Climate Change on Twitter: A Corpus-Assisted Discourse Study. In *Corpus Pragmatics* (Issue 0123456789). Springer International Publishing. <https://doi.org/10.1007/s41701-023-00140-3>
- Parvaresh, V. (2023). Covertly communicated hate speech: A corpus-assisted pragmatic study. *Journal of Pragmatics*, 205, 63–77. <https://doi.org/10.1016/j.pragma.2022.12.009>

- Pizarro, P. A. (2019). MadSex: collecting a spoken corpus of indirectly elicited sexual concepts. *Language Resources and Evaluation*, 53(1), 191–207. <https://doi.org/10.1007/s10579-018-9435-x>
- Rosli, N. N., Hamzah, N., Zaini, M. F., Baharum, H., Mohd, F. H., Jabar, N. A., Damit, A. R., & Omar, R. (2022). Sentiment analysis of emotional words in a classical text web corpus. *AIP Conference Proceedings*, 2644(November), 030027. <https://doi.org/10.1063/5.0104738>
- Smyth, C. (2016). An Introduction to Corpus Linguistics. *Bulletin of Tokyo Denki University, Arts and Sciences*, 14, 105–109. <https://doi.org/10.1177/00754240022004965>
- Szczygłowska, T. (2022). Exploring Obama's and Trump's Political Discourse through the Lens of Wordlists, Keywords and Clusters. *Brno Studies in English*, 48(1), 93–116. <https://doi.org/10.5817/BSE2022-1-5>
- Yüksel, H. G., Mercanoğlu, H. G., & Yılmaz, M. B. (2022). Digital flashcards vs. wordlists for learning technical vocabulary. *Computer Assisted Language Learning*, 35(8), 2001–2017. <https://doi.org/10.1080/09588221.2020.1854312>
- Zaini, M. F., Muhammad, M. M., Goyak, F., Saradin, A., Osman, Z., Redzwan, H. F. M., & Al Muhsin, M. A. (2022). Geometric Lexical Representative Perspectives: The Impact of Threshold Values Through #LancsBox Software. *AIP Conference Proceedings*, 2644(November). <https://doi.org/10.1063/5.0104817>
- Zaini, M. F., Sarudin, A., Muhammad, M. M., & Abu Bakar, S. S. (2020). Representatif Leksikal Ukuran sebagai Metafora Linguistik berdasarkan Teks Klasik Melayu (Representatives of Lexical Ukuran as Linguistics Metaphors Based on Malay Classic Text). *GEMA Online® Journal of Language Studies*, 20(2), 168–187. <https://doi.org/10.17576/gema-2020-2002-10>