



## **REPRESENTATION OF THE WORD REMAJA IN ONLINE MEDIA: CORPUS-ASSISTED DISCOURSE STUDY**

**Winda Ayuanda<sup>1</sup>, T. Silvana Sinar<sup>2</sup>, Mahriyuni<sup>3</sup>**

**Department of Linguistics, Graduated Program, Universitas Sumatera Utara, Medan, Indonesia<sup>123</sup>**

**windaayuanda509@gmail.com<sup>1</sup>, tengkusilvana@usu.ac.id<sup>2</sup>, mahriyuni@usu.ac.id<sup>3</sup>**

---

Accepted: December, 12<sup>nd</sup> 2024

Published: December, 26<sup>th</sup> 2024

---

### **Abstract**

*Reporting about adolescents in the mass media has a tendency towards criminal acts and delinquent behavior. The aim of this research is to look at the representation of the word 'remaja' in the online media Detikcom from 2022 to 2024 based on a corpus-assisted discourse study. This research uses a mixed approach, with the data source in the form of a corpus composed of Detikcom news articles related to adolescent issues. The corpus has a total number of 325,313 tokens and 26,652 types of words. Through collocation analysis, 45 noun, adjectival, and verbal collocates were taken (15 collocates per word class), which significantly collocated with the word remaja. There are words that explicitly show negative meanings, such as kenakalan, tawuran, gerombolan, mesum, ugal-ugalan, brutal, sadis, mabuk, and tewas. Apart from that, there are also two words that show positive meanings, such as romantis and mengatasi. From the semantic prosody analysis, it can be seen that the word remaja tends to collocate with words that have negative prosody. Based on the concordance lines that have been studied, it was found that there were many crime news that featured adolescents either as perpetrators or as victims of crime. The most frequently found semantic prosody category was negative absolute with 26 collocates, followed by the flexible extremely negative category with 13 collocates. Meanwhile, both the absolute positive and the flexible extremely positive category only have 3 collocates. It can be said that society's view of adolescents is not much different from online media, which is dominated by negative news about adolescents. This ultimately shapes adolescent discourse, which can influence how adolescents are seen and treated in society.*

**Key words:** *corpus linguistics, discourse analysis, collocation, adolescents*

**How to Cite:** Ayuanda. W., Sinar. T.S., Mahriyuni. (2024). Representation of The Word Remaja in Online Media: Corpus-Assisted Discourse Study. JUPIIS: Jurnal Pendidikan Ilmu-ilmu Sosial (164-175)

---

\*Corresponding author:  
windaayuanda509@gmail.com

ISSN 2085-482X (Print)  
ISSN 2407-7429 (Online)

## INTRODUCTION

Various issues and events that occur in adolescents 'remaja', or how the media frames the term adolescent, are able to shape the image of adolescents themselves in the public mind. Advances in technology have made it easy to record everything that happens today in online media, leaving digital traces. In this regard, it is important to study the representation of a topic—in this case, adolescents—in online media. The accessibility of news, articles, and writings relating to adolescents is very easy to reach, so images about adolescents can spread quickly. Through discourse analysis using the Corpus-Assisted Discourse Study (CADS) framework, it can be understood how adolescent group identities are formed and represented in the media. This can help identify issues and their impact on public perception.

Adolescence is a crucial phase in human life, namely a transitional phase from childhood to adulthood. As an adolescent, an individual experiences many changes both physically and emotionally, such as hormonal changes, the development of personal identity, and the search for values and beliefs. Adolescents also face increasingly complex relationships or social lives. According to (Erikson, 1968: 27), the ideological structure of the environment becomes important for the ego in adolescence because without the ideological simplification of the universe, the adolescent ego cannot organize experiences according to its specific capacities and widespread involvement, so that adolescence is a stage where the individual is closer to its historical period compared with the early stages of childhood development.

Reporting about adolescents in the mass media is perceived to have a tendency towards delinquent behavior. Fahmi & Wijayanti (2023: 341) stated that risky behavior carried out by adolescents is often reported by the mass media, both through print news and electronic news. Juvenile delinquency reported in various mass media is considered increasingly disturbing and dangerous to society, some of which include brawls, mugging, extortion, theft, robbery, abuse, drug abuse such as narcotics, free sex, and various other forms of deviant behavior (Wulandari & Hodriani, 2019: 140). Many acts of juvenile delinquency are starting to experience a shift, initially just ordinary juvenile delinquency, now tending to expand

into criminal aspects that legally violate the provisions of the criminal law (Karlina, 2020). Therefore, in this research, the corpus linguistic method is used to confirm this opinion based on more empirical data, namely language data taken from articles in the online media Detikcom.

Studies on the representation of adolescents have been carried out previously, but with data sources in the form of films and music videos, such as research by Anugrah and Suwanto (2019), who conducted research on the representation of adolescents' self-concept in the film *Lady Bird* using Roland Barthes' semiotic analysis approach. The primary data source for this research is film scenes related to teenagers' self-concept. Wulandari dan Yuliarti (2021) used Charles Sanders Peirce's semiotic approach to analyze the representation of teenage identity in the music video of the South Korean boy group NCT Dream. The objects of this research are seven music videos for the boy group NCT Dream. Rahmawati and Dewi (2023) studied the representation of adolescents in Indonesian films using the genre analysis method by Jane Stokes by identifying six genre conventions, which include setting, location, narrative events, characters, iconography, and plot structure.

In the three previous studies, the representation of adolescents in online media has not been studied. Previous studies mostly used semiotic approaches and genre analysis methods as their main analytical approaches and did not provide analysis of discursive strategies, while this research will use the CADS framework. This is a gap that will be filled by this research. The online media used in this research is Detikcom. The research object is the corpus of Detikcom online media which consists of 1,119 articles. This is a novelty and significant factor for conducting this research, to see how adolescents are described through the practice of using language with a time span from January 1, 2022, to April 2, 2024. The time span was limited by considering the novelty of the data to be analyzed, the amount of time to be spent on conducting analysis, and data from this time period is deemed sufficient to show the representation of adolescents in the online media Detikcom.

Based on this explanation, there are two research problem formulations that will make this research more focused: 1) How is the description of the word *remaja* based on the significant collocates that appear? 2) What is the semantic prosody of words that collocate

with the word *remaja*?

Discourse is a source of knowledge in the form of people's generalizations about language which are made based on the discourse they follow and the results of knowledge, namely the application of what they already know in creating and interpreting new discourse (Johnstone, 2008: 3). Discourse analysis is research that examines how language in discourse is able to form meaning and influence the minds of readers or listeners. Discourse analysts are interested in levels of frequency that achieve significance in the context of perception, so that regularities in discourse are linguistic features that appear in a particular environment with significant frequency, and will usually adopt the methodology of the descriptive linguistic tradition to describe the linguistic forms contained in their data (Brown & Yule, 1988: 22).

Corpus linguistics and critical discourse analysis have lived separate lives, but ultimately they can work well together in revealing how language 'works' in social and not just structural contexts. According to Mautner, the benefit of accessing large computer-managed corpora lies in increasing the empirical confidence of the analysis, which can serve to counter the accusation, often leveled at critical discourse analysis, that particular texts are selected according to the researcher's political agenda (Mautner, 2009: 32). By using a corpus, at least cognitive biases can be limited because it is difficult to be selective about one newspaper article when looking at hundreds of articles, and research starts from a position where the data itself has not been selected to confirm consciously or unconsciously existing biases (Baker, 2006: 12).

Corpus-based discourse analysis is generally also called corpus-assisted discourse study (Eriyanto, 2022). Basically, the CADS method is a method for describing and expressing discourse with the help of a corpus (Mautner, 1995). CADS is a combination of two methods, namely critical discourse analysis and corpus linguistics. CADS sees language as an ideological practice, and the choice of words or language critically depicts inequality, discrimination, and so on. The tools and concepts used in CADS are frequency, collocation, semantic preference, and semantic prosody, thereby producing case studies of

ideology-laden lexis as well as developing the analytical concepts necessary to extract socially important information from corpora (Mautner, 2009).

Previous studies that used the CADS method include Baker et al., (2013) who analyzed representation patterns around the word 'Muslim'. From the analysis of noun collocations, the category 'conflict' has quite a lot of lexical content. This research provides insight into the categorization of semantic preferences and the analysis of the most frequently occurring collocations. Research on five identical Indonesian language lemmas (*hasil, efek, konsekuensi, dampak* and *akibat*) by Prihantoro, (2015) also contributes to insights into semantic prosody for current research. From this research it can be learned that prosody is divided into positive, negative and neutral prosody. In addition, prosody has flexible and absolute properties.

The next corpus-based discourse analysis research was carried out by Al Fajri (2017) who analyzed the discourse around the word immigrant in the 'ukWac' (Web as Corpus) corpus. Concordance analysis in this research was applied to examine the data more qualitatively. This research also uses a span or collocation span of three words on both sides of the word node, using log-likelihood statistics in collocation analysis. The most recent relevant research is by Santyarini and Fajri (2024) which analyzes noun and verb collocations in diabetes discourse. This study emphasizes the importance of achieving more balanced diabetes coverage in the news media in Indonesia. Integrating depictions that highlight structural issues into news narratives is critical to fostering comprehensive understanding, gaining public support for policy changes at the community level, and reducing the stigma associated with diabetes sufferers.

## METHODOLOGY

This research used a mixed-approach design, namely a combination of quantitative and qualitative approaches. In the quantitative design, an analysis was carried out on the frequency of occurrence of words that most frequently collocate with the word 'remaja' and an investigation into significant collocations. In the qualitative design, concordance line analysis was carried out to determine semantic prosody in the collocated word data, and then interpretation of the collocates was carried out.

The data source for this research was a corpus built from a collection of Detikcom article texts from 2022 to 2024. The number of news articles collected was 1,073 article titles. This corpus has a total number of tokens of 325,313 tokens and word types of 26,652 types. The data obtained in this research were words that collocate with the word adolescent. The limitations of research data sources are data from noun, verb, and adjectival word classes.

Detikcom's online media corpus was created by scraping data via Google Colab with the Python programming language. The instruments for this research were Google Colab and the CQPweb corpus tool. Google Colab or Google Colaboratory is a cloud-based platform (delivery of computing services over the internet) that allows users to write and run Python code via a web browser. CQPweb is a web-based corpus analysis system with analysis options including concordance, collocation, distribution tables and graphs, frequency lists, and keywords (Hardie, 2012: 380).

The data of collocation was carried out using the CQPweb corpus tool. The range or collocation window used in this research was three words from both the right and left sides of the word node. The statistic used in collocation searches in this study was log-likelihood (LL). In this research, the filters used were three word class codes or POS tags, namely: NN (POS tag for nouns), JJ (POS tag for adjectives), and VB (POS tag for verbs). Collocate data was limited by considering the data with the highest log-likelihood score. The collocate data taken consisted of 15 noun, 15 adjectival, and 15 verba collocates. To find the semantic prosody of words that collocate with the word *remaja*, a concordance analysis was carried out by looking at the left and right contexts of the node word.

## RESULT AND DISCUSSION

One of the CADS analysis techniques used in this research was collocation analysis. The collocations of the word *remaja* focused on the noun, adjectival, and verb word classes. The words that often appear together or were collocated in the text can provide clues about the meaning and relationships between words, so they can show how something was represented. As the word *brawl* often appears together with the word *teenager*, it can shape the public's perspective on negative images of

teenagers. After searching for collocations in the corpus, 157 noun collocates, 20 adjective collocates and 134 verb collocates were found. This study only took the collocates with the highest log-likelihood scores for data effectiveness. The following is an explanation of each collocate based on its word class.

### 1. Noun Collocate

In accordance with the predetermined limits, the noun collocates taken are the top 15 collocates, namely those with the highest log-likelihood scores. From the list of noun collocates that appear, further observations are made to see whether all the words that appear are words in the noun word class, because errors may occur when the corpus tool reads the data. A list of noun collocates from the word *remaja* can be seen in Table 1.

**Table 1. Noun collocate of *remaja***

No	Word (NN)	Observed collocate frequency	Log-likelihood
1	Perempuan	172	482.739
2	Kenakalan	86	402.143
3	Putri	88	332.077
4	Laki-laki	84	238.217
5	Kelompok	142	213.913
6	Asal	48	86.461
7	Tawuran	128	62.513
8	Tahun	120	50.331
9	Kalangan	19	49.935
10	Usia	38	47.140
11	Aksi	97	47.116
12	Gerombolan	19	46.317
13	Kamar tidur	9	31.180
14	Anak-anak	46	29.177
15	Perwira	7	22.451

The noun collocate was the most common collocate that appeared. From the data in the table above, the noun collocate with the highest log-likelihood score was the word *perempuan* 'woman' with a score of 482,739. This showed that adolescents of the female gender were reported more often than those of the male gender. From the data found, the frequency of occurrence of the word *perempuan* was 172 times, while the word *laki-laki* 'man' was 84 times. The appearance of the word *putri* 'woman' also has high significance, with a frequency of appearance of 88 times,

which once again showed that women were more likely to be reported in online media than men.

There were also words that literally have a negative meaning in the data, namely the word *tawuran* 'brawl' which appeared 128 times. This figure was quite high, even though the log-likelihood score was ranked 7th, namely 62,513. A brawl is a fight carried out in groups. Brawls were identically carried out by teenagers. The words *kelompok* and *gerombolan* in the table above also showed that adolescents often carry out activities in groups.

Apart from the word *tawuran*, the word that clearly showed a negative meaning was the word *kenakalan* 'delinquency' which appeared 86 times. In contrast to these two words, the word *aksi* 'action' could actually have positive and negative meanings. As in the KBBI (Kamus Besar Bahasa Indonesia), the word *aksi* means 'movement' (for example an action to collect funds) and 'action' (for example an action of retaliation). To see whether in practice the use of this word in a teenage context shows a positive or negative meaning, it will be discussed further in the semantic prosody section.

## 2. Adjectival Collocate

Adjectival collocates are the least frequently found collocates. The data analyzed was also limited to 15 collocates that had the highest log-likelihood scores. From the list of adjectival collocates that appeared, further observations were made to see whether all the words that appeared were words in the adjective word class or not. A list of adjectival collocates from the word *remaja* can be seen in Table 2.

**Table 2. Adjectival collocate of *remaja***

No	Word (NN)	Observed collocate frequency	Log-likelihood
1	Jompo	9	46.604
2	Romantis	5	15.774
3	Rentan	9	13.801
4	Hitam	15	13.196
5	Mesum	7	9.408
6	Ugal-ugalan	5	9.046
7	Brutal	5	7.733
8	Misterius	5	7.166
9	Sadis	5	6.175
10	Sadar	5	4.962

11	Nekat	7	4.300
12	Positif	11	4.259
13	Khusus	6	3.990
14	Mabuk	6	3.751
15	Sosial	34	1.891

From the table above, it could be seen that the word *jompo* 'elderly' has the highest log-likelihood score, namely 46,604, even though its frequency of occurrence was only 9 times. The term *remaja jompo* is widely discussed, both on social media and in the real world. The phenomenon of *remaja jompo* has been researched by Sujibto (2022). He explained that the discourse on *remaja jompo* was born from the influence of teenagers' lifestyles under health protocols during the COVID-19 pandemic. Schools and lectures that were conducted online, as well as the habit of being lazy due to exposure to gadgets and technology, have changed the mental and physical health conditions of teenagers until, finally, *remaja jompo* has become a social phenomenon among young people.

From the list of adjectival collocates, there were words that explicitly showed negative meanings, such as *rentan* 'vulnerable', *mesum* 'perverted', *ugal-ugalan* 'reckless', *brutal* 'brutal', *misterius* 'mysterious', *sadis* 'sadistic', *nekat* 'reckless', and *mabuk* 'drunk'. Meanwhile, other words still seem to have a neutral meaning. To see whether in practice the use of those words in a teenage context shows a positive or negative meaning, it will be discussed further in the semantic prosody section.

## 3. Verbal Collocates

As with the data analyzed for noun and adjectival collocates previously, verbal collocate data was also taken based on the highest log-likelihood scores, namely the top 15 ranked collocates. Before the data was presented, observations were made first to see whether all the words that appeared were words belonging to the verb class. A list of verbal collocates from the word *remaja* can be seen in Table 3.

**Table 3. Verbal collocate of *remaja***

No	Word (NN)	Observed collocate frequency	Log-likelihood
1	Berinisial	179	389.511
2	Berusia	141	323.600

3	Mengamankan	130	251.640
4	Bernama	61	103.351
5	Menangkap	56	95.652
6	Terlibat	71	91.420
7	Terlihat	58	83.855
8	Berboncengan	25	56.303
9	Tampak	38	54.602
10	Tewas	58	50.603
11	Membubarkan	17	41.203
12	Menimpa	15	37.229
13	Mengatasi	15	33.978
14	Berbaju	11	29.456
15	Berkelahi	18	28.201

Verbal collocates data showed agreement in the frequency of occurrence and the highest log-likelihood score, which was found in words *berinisial* 'with initials'. The word *berinisial* appeared 179 times with a log-likelihood score of 389,511. The word *berinisial* always refers to the victim or perpetrator of a crime, so it can be said that the appearance of the word *berinisial* described the life of adolescents who were not far from committing a crime. The word *berinisial* has the function of implicitly indicating the identity of the victim or perpetrator, just as the word *berbaju* 'wearing clothes' in number 14 also functions to show the identity or characteristics of the perpetrator or victim without mentioning their real name.

The lives of adolescents who are not far from crime were strengthened by the appearance of the word *tewas* 'dead' in number 10, with a fairly high frequency of occurrence, namely 58 times. In the KBBI, the word *tewas* means dead (in war, disaster, etc.). After further investigation through concordance analysis, apart from referring to the tragedy of traffic accidents, drowning, and disease, the word *tewas* was also related to deaths caused by murder, beatings, and sharp weapons, which reinforces the picture of criminal

behavior that often occurs among adolescents.

Other words such as *berusia* 'aged', *terlihat* 'seen', *mengamankan* 'secured', and *mengatasi* 'overcome', intuitively suggest neutral meanings. This means that these words can have positive or negative meanings. To see whether in practice the use of those words in a teenage context shows a positive or negative meaning, it will be discussed further in the semantic prosody section.

#### 4. Semantic Prosody

In this section, it will be explained the semantic prosody of words that collocate with the word *remaja*, so that from this analysis it could also be seen whether the tendency of the word *remaja* was in a positive or negative polarity. As per Prihantoro's (2015) research, it was stated that apart from polarity (positive, negative, or neutral), another characteristic of prosody is 'flexible' or 'absolute'. In contrast to 'absolute', words included in the 'flexible' category are not definite in prosody, so further analysis must be carried out through collocation by looking at the context around the word, because collocation plays an important role in determining prosody. This was also conveyed by Hauser and Schwarz (2023: 98), even words that seem neutral can cause positive or negative responses. This applies to words that mostly appear together with positive (or negative) words in natural language.

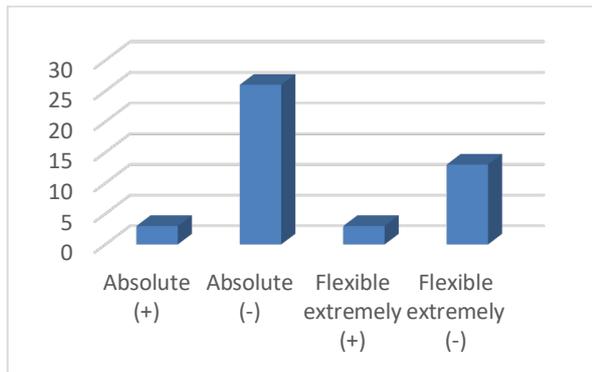
Concordance analysis was carried out based on the concordance lines that appeared when the word/token collocated with the word *remaja* in the corpus. Therefore, the word *perempuan*, which in the results of this analysis was included in flexible prosody with an extremely negative tendency (see Table 4), may be different if it was related to a discourse context other than adolescents. The total data categorization was visualized through the diagram in Figure 1.

**Table 4. Prosodic categorization of *remaja*'s collocations**

No	Token	Prosody		Absolute	Flexible	
		(+)	(-)		extremely (+)	extremely (-)
1	Perempuan	✓	✓	×	×	✓
2	Kenakalan	×	✓	✓	×	×
3	Putri	✓	✓	×	×	✓
4	Laki-laki	✓	✓	×	×	✓
5	Kelompok	✓	✓	×	×	✓

6	Asal	✓	✓	x	x	✓
7	Tawuran	x	✓	✓	x	x
8	Tahun	✓	✓	x	x	✓
9	Kalangan	✓	✓	x	x	✓
10	Usia	✓	✓	x	x	✓
11	Aksi	x	✓	✓	x	x
12	Gerombolan	x	✓	✓	x	x
13	Kamar tidur	✓	✓	x	✓	x
14	Anak-anak	✓	✓	x	x	✓
15	Perwira	✓	x	✓	x	x
16	Jompo	✓	✓	x	✓	x
17	Romantis	✓	x	✓	x	x
18	Rentan	x	✓	✓	x	x
19	Hitam	x	✓	✓	x	x
20	Mesum	x	✓	✓	x	x
21	Ugal-ugalan	x	✓	✓	x	x
22	Brutal	x	✓	✓	x	x
23	Misterius	x	✓	✓	x	x
24	Sadis	x	✓	✓	x	x
25	Sadar	✓	✓	x	✓	x
26	Nekat	x	✓	✓	x	x
27	Positif	x	✓	✓	x	x
28	Khusus	✓	✓	x	x	✓
29	Mabuk	x	✓	✓	x	x
30	Sosial	✓	✓	x	x	✓
31	Berinisial	x	✓	✓	x	x
32	Berusia	✓	✓	x	x	✓
33	Mengamankan	x	✓	✓	x	x
34	Bernama	✓	✓	x	x	✓
35	Menangkap	x	✓	✓	x	x
36	Terlibat	x	✓	✓	x	x
37	Terlihat	x	✓	✓	x	x
38	Berboncengan	x	✓	✓	x	x
39	Tampak	x	✓	✓	x	x
40	Tewas	x	✓	✓	x	x

4	Membubarkan	x	✓	✓	x	x
4	Menimpa	x	✓	✓	x	x
4	Mengatasi	✓	x	✓	x	x
4	Berbaju	x	✓	✓	x	x
4	Berkelahi	x	✓	✓	x	x



**Figure 1. Prosodic categorization of remaja's collocations**

From Figure 1, it could be seen that the word *remaja* tended to collocate with words that have negative prosody. Based on the concordance lines that have been researched, it was found that there were many crime news stories that featured adolescents either as perpetrators or as victims of crime. Very little news about positive things, such as the achievements or success of adolescents, appeared in the online media Detikcom, so their significance was low in the corpus. This was in line with the teenage phenomenon, which was previously explained in the background, namely that adolescents are identical with delinquent behavior.

The findings show that the prosody category with the highest number of collocates was absolute negative with 26 collocates, followed by the flexible extremely negative category with 13 collocates. Meanwhile, both the absolute positive and the flexible extremely positive category only have 3 collocates. The following will explain the four categories of prosody by showing examples of collocates and samples of concordance lines in each category.

#### **a. Absolute (+)**

This category contained words that consistently displayed positive prosody throughout their concordance lines. In the data above, there were only two words that are categorized as absolute positive prosody,

namely the words *romantis* 'romantic', *perwira* and *mengatasi* 'overcoming'. This number made this category the smallest category among the other categories. The following is an explanation of the concordance analysis.

N	Concordance		
1	Film <b>romantis</b> barat	remaja	menjadi salah satu genre populer di incaran penonton .
2	Itu tadi beberapa judul Film <b>romantis</b> barat	remaja	terbaik yang bisa detikers tonton .
3	The Kissing Booth tentunya masuk dalam daftar film <b>romantis</b> barat	remaja	. Film Netflix ini dibintangi oleh Joey King ,

The collocate *romantis* was used to describe the adjective of the word film in lines (1), (2), and (3), which indicates a positive meaning. In this context, the word node *remaja* and the collocate *romantis* were used to denote entertainment that can be enjoyed by the public. This made the concordance lines (1), (2), and (3) have positive prosody. Additionally, the words *populer* 'popular' and *terbaik* 'best' in lines (1) and (2) strengthened the positive meaning of these concordance lines.

#### **b. Absolute (-)**

This category contained words that consistently displayed negative prosody throughout their concordance lines. In the data above, there were 26 words in the absolute negative prosody category. This number made this category the largest among other categories. Several words in this category have negative lexical meanings, such as *kenakalan*, *tawuran*, *mesum*, *mabuk*, *brutal*, and others. There were also several words that still sounded neutral in meaning, such as the words *berinisial*, *positif*, *mengamankan*, *berboncengan*, *menimpa*, but after further

research, these words were included in the absolute negative category. The following is a presentation regarding the concordance analysis of several of these words.

N o	Concordance		
1	Nasib sial <b>menimpa</b> dua	remaj a	berinisial AH ( 21 ) dan AI ( 21 )
2	Dijelaskan , kronologi penjambretan yang <b>menimpa</b>	remaj a	asal Banjar Dinas Musi ini bermula ketika korban hendak pulang.
3	Ia menerangkan , aksi kekerasan yang <b>menimpa</b> seorang	remaj a	perempuan itu terjadi di Kecamatan Babakan , Kabupaten Cirebon ,
4	Polisi telah melakukan penyelidikan terkait pemukulan yang <b>menimpa</b> dua	remaj a	berusia 14 dan 16 tahun di Grogol Petamburan , Jakarta

The collocate *menimpa* 'afflict' in the concordance sample above appeared after words with negative meanings such as *nasib sial* 'bad luck' in line (1), *penjambretan* 'mugging' in line (2), *kekerasan* 'violence' in line (3), and *pemukulan* 'beating' in line (4). This caused the word *menimpa*, which was collocated with the word *remaja*, to have a negative prosody. In this context, adolescents were depicted as victims of criminal acts.

N o	Concordance		
1	Hasil pemeriksaan , tiga	remaj a	tersebut <b>positif</b> mengonsumsi ganja.
2	Saat ini disebutkan semakin banyak	remaj a	yang <b>positif</b> HIV akibat perilaku hubungan seks berisiko.
3	"Karena tidak ada barang bukti	remaj a	yang <b>positif</b> psikotropika itu kita serahkan ke BNN untuk rehabilitasi
4	Ikatan Dokter Anak Indonesia ( IDAI ) menyebutkan penyebab	remaj a	<b>positif</b> HIV karena narkoba hingga hubungan seks berisiko.

The concordance lines that contained the collocate *positif* 'positive' above showed that the practice of using the word *positif* in the word *remaja* depicted negative things. The word *positif* appeared alongside the words *ganja* 'marijuana', *HIV*, and *psikotropika* 'psychotropics'. This resulted in the four lines of the concordance having negative prosody.

The word *berisiko* 'risk' in lines (2) and (4) strengthened the negative meaning in the concordance lines.

### c. Flexible Extremely (+)

This category contained words that not only displayed positive prosody but also displayed negative prosody in their concordance lines. However, there was a tendency for extremely positive prosody in this category. From the previous data, there are three words that fall into the flexible extremely positive prosody category, namely *sadar* 'conscious', *jompo* 'decrepit', and *kamar tidur* 'bedroom'. The following is a presentation regarding the concordance analysis of one of the words.

N o	Concordance		
1	dan menyediakan pilihan yang lebih sehat , secara langsung membuat	remaj a	lebih <b>sadar</b> akan pilihan mandiri mereka .
2	Meski mengalami luka bakar serius , kondisi	remaj a	itu masih <b>sadar</b> dan mendapat perawatan intensif .
3	Pada fase ini , secara tidak <b>sadar</b>	remaj a	merasakan gangguan kecemasan yang bisa mempengaruhi kesejahteraan mentalnya.

Concordance lines (1) and (2) had positive prosody. In line (1), the collocate *sadar* appeared together with the word *pilihan mandiri* 'independent choice'. After analyzing the broader context, this illustrated the character of adolescents in making independent choices that have the potential to form a healthier lifestyle. The collocate *sadar* and the word node *remaja* in line (2) were used to show the good side of an event, as indicated by the words *masih sadar* and *mendapat perawatan intensif* 'receiving intensive care'. Therefore, the word *sadar* in both lines had positive prosody.

In contrast to lines (1) and (2), line (3) actually gave rise to negative prosody. The collocate *sadar* in this line began with the negation of *tidak* 'not', to describe how anxiety disorders can occur in adolescents. From this explanation, it can be seen that collocate *sadar* and the word node *remaja* could have positive and negative prosody, but the frequency of

their occurrence in the corpus tended to display positive prosody.

#### d. Flexible Extremely (-)

This category contained words that not only displayed positive prosody, but also displayed negative prosody in their concordance lines. However, there was a tendency for extremely negative prosody in this category. From previous data, there were 12 words that fall into the category of flexible extremely negative prosody, such as the words *perempuan*, *sekelompok* 'a group', *laki-laki*, *sepasang* 'pair', etc. The following is an explanation of the concordance analysis of one of these words.

N	Concordance		
1	pencabulan , dilaporkan ke Polres Metro Bekasi Kota . Korban	remaj a	<b>perempuan</b> usia 17 tahun , " kata Kepala Bidang Humas
2	Sebuah video mempertontonk an dua	remaj a	<b>perempuan</b> berkelahi di tengah jalan . Video perkelahian di salah
3	Polisi tengah menyelidiki kasus kekerasan yang menimpa	remaj a	<b>perempuan</b> di Kabupaten Cirebon . Korban dianiaya oleh pelaku yang
4	Irjen Iqbal menyoroti tanggung jawab khusus yang dimiliki oleh Bintara	Remaj a	<b>Perempuan</b> mengingat mereka menjadi idola .

The collocate *perempuan* in the four lines of the concordance above gave the character of gender identity in adolescents. Concordance lines (1), (2), and (3) have negative prosody. As can be seen in lines (1) and (3), the collocate *perempuan* and the word node *remaja* were used to describe adolescents as victims of sexual abuse and violence. Meanwhile, in line (2), the collocate *perempuan* and the word node *remaja* showed adolescents as perpetrators of delinquent acts, this was marked by the appearance of the word *berkelahi* 'fighting'.

In contrast to the three previous lines of the concordance, line (4) contained a positive meaning. The association of the word node *remaja* with the collocate *perempuan* in line (4) was used to show that adolescents can become idols, especially when they become

*Bintara Remaja Perempuan*. Line (4) had better context and contained positive prosody.

N	Concordance		
1	Dari hasil pemeriksaan , petugas mengidentifikasi korban merupakan	remaj a	<b>laki-laki</b> berusia 13 tahun . Korban diketahui berinisial MR .
2	Polisi mengamankan 24	remaj a	<b>laki-laki</b> hendak tawuran di Pintu Tol Desari , Depok
3	Ary awalnya diringkus polisi karena diduga menyodomi seorang	remaj a	<b>laki-laki</b> berusia 16 tahun di kamar hotel kelas melati di
4	memiliki tingkat yang berbeda berdasarkan usia . Hasilnya menunjukkan bahwa	remaj a	<b>laki-laki</b> mengalami peningkatan yang lebih besar dibandingkan remaja <b>perempuan</b> .

The collocate *laki-laki* and the word node *remaja* in lines (1), (2), and (3) have negative prosody. The association of the word node *remaja* with the collocate *laki-laki* in lines (1) and (3) described adolescents as victims of violence and abuse. Line (1) in the concordance data above did not yet show what action made the teenager a victim, but after analyzing the broader context, the teenage boy was a victim who was found on the edge of an irrigation canal and was suspected of experiencing violence before being found. This caused the prosody in lines (1) and (3) to be negative prosody.

On the other hand, the description of adolescents in line (2) had a contradictory meaning. Adolescents were described as perpetrators of crimes in line (2), with the appearance of the collocate *laki-laki* followed by the word *tawuran*. As is known, *tawuran* is deviant behavior carried out by a group of people, so this concordance line showed negative prosody. Different from the three lines that have been explained, line (4) had a positive prosody, marked by the appearance of the word *peningkatan* 'enhancement' after the words node *remaja* and the collocate *laki-laki*. In this context, the word *peningkatan* caused the death rate for adolescent boys to decrease more than for girls. Therefore, it can be said that line (4) had positive prosody.

From these explanations, it can be seen that collocates *perempuan* and *laki-laki* for the

word node *remaja* can have negative and positive prosody, but the frequency of their occurrence in the corpus tends to show negative prosody.

## CONCLUSION

After exploring for collocations in the corpus using the CQPweb corpus tool, 157 noun collocates, 20 adjectival collocates, and 134 verbal collocates were found. From the collocation analysis, it can be seen how adolescents were represented through certain word choices and combinations in online media. From the list of noun, adjectival, and verbal collocates, there were words that explicitly show negative meanings, such as *kenakalan, tawuran, gerombolan, mesum, ugal-ugalan, brutal, sadis, mabuk, tewas, and berkelahi*. There were also two words that showed positive meanings, such as *romantis, perwira, and mengatasi*. Meanwhile, the other words still look neutral.

Then, from the semantic prosody analysis, it can be seen that the word *remaja* tends to collocate with words that have negative prosody. Based on the concordance lines that have been studied, it was found that there were many crime news that featured adolescents either as perpetrators or as victims of crime. Very little news about positive things, such as the achievements or success of adolescents, appeared in the Detikcom online media, so their significance was low in the corpus. This was in line with the teenage phenomenon, which was previously explained in the background, namely that adolescents are identical with delinquent behavior. The findings showed that the prosody category with the highest number of collocates was absolute negative with 26 collocates, followed by the flexible extremely negative category with 13 collocates. It can be said that society's view of adolescents is not much different from online media, which is dominated by negative news about adolescents. This ultimately shapes adolescent discourse, which can influence how adolescents are seen and treated in society.

By using corpus linguistic methodology, the analysis of adolescent discourse in online media provides several novelties. Firstly, it gives more objective results because they use more empirical data compared to conventional discourse analysis methods. This research also reveals how media shapes social perceptions of adolescents

through language aspects, and the data in this study is able to provide the latest representation of adolescent discourse.

## REFERENCE LIST

- Al Fajri, M. S. (2017). Hegemonic and Minority Discourses around Immigrants: A Corpus-Based Critical Discourse Analysis. *Indonesian Journal of Applied Linguistics*, 7(2), 143. <https://doi.org/10.17509/ijal.v7i2.8349>
- Anugrah, I., & Suwanto, D. H. (2019). Representasi Konsep Diri Remaja pada Film *Lady Bird* (Analisis Semiotika Roland Barthes). *Lektur: Jurnal Ilmu Komunikasi*, 2(3). <https://doi.org/10.21831/lektur.v2i3.16313>
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: 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>
- Brown, G., & Yule, G. (1988). *Discourse Analysis*. Cambridge: Cambridge University Press.
- Erikson, E. H. (1968). *Identity: Youth and Crisis*. New York: W. W. Norton.
- Eriyanto. (2022). *Analisis Wacana Kritis Berbasis Korpus (Corpus-Assisted Discourse Studies/CADS) Pengantar Metode Analisis Teks Media*. Bandung: PT. Remaja Rosdakarya.
- Fahmi, A. M., & Wijayanti, Q. N. (2023). Analisis Peran Komunikasi antar Pribadi Orangtua dengan Anak dalam Mencegah Kenakalan di Era Zaman Sekarang. *Jurnal Media Akademik*, 1(1), 338-347.
- Hardie, A. (2012). CQPweb — combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380-409. <https://doi.org/10.1075/ijcl.17.3.04har>
- Hauser, D. J., & Schwarz, N. (2023). Semantic Prosody: How Neutral Words With Collocational Positivity/Negativity Color Evaluative Judgments. *Current Directions in Psychological Science*, 32(2), 98-104. <https://doi.org/10.1177/09637214221127978>
- Johnstone, B. (2008). *Discourse Analysis*. Oxford: Blackwell Publishing Ltd.
- Karlina, L. (2020). Fenomena Terjadinya

- Kenakalan Remaja. *Jurnal Edukasi Nonformal*, 1(1). 147-158.
- Mautner, G. (2009). Corpora and Critical Discourse Analysis. In P. Baker (Ed.), *Contemporary Corpus Linguistics* (pp. 32–46). Continuum International Publishing Group.
- Prihantoro, P. (2015). Semantic Prosody of Words of Effects In Indonesian. *Indonesian Journal of Applied Linguistics*, 5(1), 101. <https://doi.org/10.17509/ijal.v5i1.837>
- Rahmawati, D., & Dewi, P. A. R. (2023). Representasi Remaja dalam Film Indonesia. *The Commercium*, 7(3), 97–110. <https://doi.org/10.26740/tc.v7i3.56987>
- Santyarini, R., & Fajri, M. S. Al. (2024). The portrayal of diabetes in the Indonesian online press: a corpus-based discourse study. *Cogent Arts & Humanities*, 11(1). <https://doi.org/10.1080/23311983.2024.2346374>
- Sujibto, B. (2022). Remaja Jompo: Diskursus dan Praktik Gaya Hidup Pandemi. *Jurnal Studi Pemuda*, 11(1), 14. <https://doi.org/10.22146/studipemudau gm.75381>
- Wulandari, D. O., & Hodriani, H. (2019). Peran Guru Pendidikan Kewarganegaraan dalam Mencegah Kenakalan Remaja di Sekolah. *Journal of Education, Humaniora and Social Sciences (JEHSS)*, 1(3), 139–147. <https://doi.org/10.34007/jehss.v1i3.28>
- Wulandari, M. A. D., & Yuliarti, M. S. (2021). Representasi Identitas Remaja dalam Video Musik Boy Group Korea Selatan “Nct Dream” (Analisis Semiotika Charles Sanders Peirce terhadap Video Musik Boy Group Korea Selatan “Nct Dream” Di Youtube). *Jurnal Komunika Massa*, 1(1).