

# Attitude Analysis Surrounding Gender-Related Words in Lelaki Harimau

Lalitia Apsari

**Abstract**—This corpus-based study is a practice of ATTITUDE analysis as an attempt to detect any inclination regarding themes surrounding gender representation in literature. The focus includes gender-related words *gadis*, *perempuan* (female), and *lelaki* (male) in *Lelaki Harimau*, a fictional novel by Eka Kurniawan. A combination of rule-based computational methods and manual approaches are used to detect and analyse ATTITUDE-conveying terms which may construct gender representation and identity. Initial evaluation shows that references to aesthetic principles and emotion have more interactions with female gender words compared to male gender words. In addition, the equally strong existence of *gadis* and *perempuan* for female gender displays interesting characterisations within a gender, suggesting that each word may serve a rather significantly different purpose. This study can be scaled to examine gender representation using notably bigger corpora and is also a preliminary case study in the development of a comprehensive and annotated open dictionary of Indonesian language together with Data Science Indonesia Foundation that is congruent for computational analysis.

**Index Terms**—Appraisal theory, ATTITUDE, corpus linguistics, gender representation.

## I. INTRODUCTION

How genders are represented in the production and reception of text has been a long and continuous debate, specifically in literary criticism. This study utilised quantitative analysis supported with qualitative detailings to examine words representing genders in Eka Kurniawan's fictional crime novel *Lelaki Harimau* (international title: *Man Tiger*). The text is analysed to detect the inclination on how femininity and masculinity are discursively performed in the text, with certain limitations. The study focuses only on 3 gender-related tokens: *gadis* and *perempuan* for female gender and *lelaki* for male gender. This paper does not utilise the characters for the analysis nor impose any generalisation on male and female characters, but rather seeks to determine how those tokens are used, what surrounding words are recognised, and whether there are certain preliminary patterns visible for further studies on femininity and masculinity in literature. Appraisal Theory is used as a basis methodology of detailed analysis to convey potential instances or patterns particularly related to ATTITUDE.

Manuscript received June 15, 2017; revised September 10, 2017.

Lalitia Apsari is with the Descriptive Linguistic Department, Faculty of Cultural Studies, University of Indonesia, Indonesia (e-mail: lalitia.apsari@ui.ac.id).

### A. Using Corpora to Analyse Gender

A corpus can be defined as a systematic collection of naturally occurring texts (of both written and spoken language). Therefore, corpus linguistics is viewed by some as an empirical method of linguistic analysis and description, using real-life examples of language data stored in corpora as the starting point [1], [2]. For many kinds of linguistic analyses, the most important method is to search the corpus in question for the occurrence of certain strings (i.e., words or phrases). As output, the software then usually provides information on the number of these strings occurring in the corpus, on the part of the corpus and/or text in which they were found, and on the (so-called) concordance-lines, which show the strings in question in context. A concordance is 'a collection of the occurrences of a word form, each in its own textual environment' [3], where textual environment equates to the immediate co-text on either side of the search item. It is also called KWIC (Key-Word In Context) analysis. This study uses the KWIC method as a significant initial component of the analysis. It is particularly useful for both hypothesis testing and hypothesis generation, which are the main purposes of the study.

### B. Importance of This Study

The availability of digital linguistic data and other rapidly growing digital sources intended to encourage and prompt research and discussions regarding public, personal, or social issues, has forged demands for development of a quantitative method for analysing people's perspectives, attitudes, and preferences. This study is a practice using *Lelaki Harimau*, a fiction novel by Eka Kurniawan, which aims to propose a preliminary phase of systematic and replicable linguistic analytical framework to analyse gender representation and identity construction in literature, supported by both quantitative and qualitative methods utilising a combination of systemic functional linguistics framework [4], Appraisal Theory, particularly the ATTITUDE types [5], and the Lövheim cube of emotion [6] to support the AFFECT analysis in the ATTITUDE types.

This study hopes to construe discerning semantic themes surrounding gender-related words *gadis*, *perempuan*, and *lelaki*, which potentially illustrate the author's perspective in defining one type of gender identity in opposition to another. Potential development aims to expand the analysis using bigger corpora to investigate the inclination in theme (beyond gender) surrounding any designated word. This study is also a part of the preliminary phase of case studies in developing a comprehensive and annotated open dictionary of Indonesian language that is compatible and congruent with libraries in any programming languages to

advance analyses of linguistic data in the Indonesian language. The project is conducted under the Data Science Indonesia Foundation.

## II. METHODOLOGY

Eka Kurniawan was born on November 28, 1975, and he cites Pramoedya Ananta Toer as one of his favorite writers. The two writers have been compared several times by critics even though the differences of their works are quite noticeable. Kurniawan's works are known to be digressive yet riveting, an ode to Indonesia's oral traditions. His writings contain relatively little dialogue and delve more into the narrative style, where he elaborates upon what the characters do and feel, and thus he is considered to be one of the strongest contemporary storytellers in Indonesia.

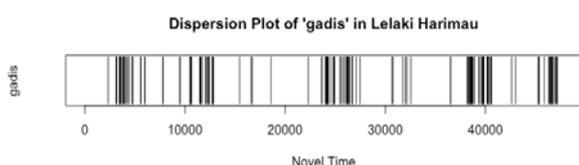
### A. Material

The methodology involves the novel *Lelaki Harimau* [7], a corpus of 42,165 tokens that requires the use of a computerised corpus analysis method, small-scale computational linguistics method, and manual analysis to detect patterns within the linguistic evidence. This corpus has a Type-Token Ratio (TTR) of 0.1264.

To ensure a more focused purpose of this study, not all gender representations are analysed. This study attempts to conduct an initial analysis of gender representation and identities in *Lelaki Harimau* using 3 selected words: *gadis* and *perempuan* for female gender and *lelaki* for male gender. These gender-related words are selected due to their positions in the top 30 of the sorted word frequency list<sup>1</sup>. Interestingly, Kurniawan doesn't use *pria* (male) or *wanita* (female), which are also terms for gender.

Initially, a Term-Document Matrix (TDM) consists of a reference corpus containing 20 Indonesian fictional novels from the year 2000 to 2017 applying Term Frequency-Inversed Document Frequency (TF-IDF) weighting is created to analyse how frequently those words appear in the reference corpus, but there are no significant differences between the three words (the TF-IDF value of *gadis*, *perempuan*, and *lelaki* in the reference corpus are 5.57, 7.21, and 4.32 respectively), and they are not significant in the reference corpus as opposed to the specific corpus used to conduct the analysis.

This study uses R software (version 3.3.3) [8] for the quantitative analyses and visualisations. Association measurements for detecting collocations are based on the likelihood ratio [9] as implemented in the 'quanteda' package [10] in R. The network graph visualisation to illustrate the interaction and networking analysis between the gender-related words and their surrounding words uses Kamada-Kawai's algorithm [11].



<sup>1</sup> A list of stop words in Indonesian is used in the preprocessing phase to exclude determiners, coordinating conjunctions, and prepositions from the word frequency list.

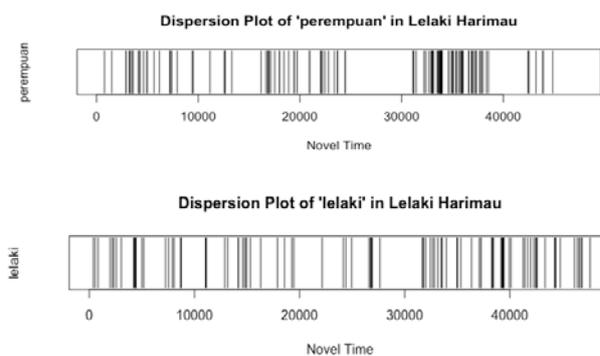


Fig. 1. Dispersion plots of *gadis*, *perempuan*, and *lelaki* in *Lelaki Harimau*.

The dispersion of *gadis*, *perempuan*, and *lelaki* are different throughout the novel as presented in Fig. 1. There is an evident cluster for *gadis* in the last chapter and one for *perempuan* in the fourth chapter, whereas the use of *lelaki* spreads with an increased frequency towards the last chapter. An interesting pattern is visible in Fig. 2 particularly in the last two chapters, where the usage frequencies of *gadis* and *perempuan* in both chapters are in opposition.

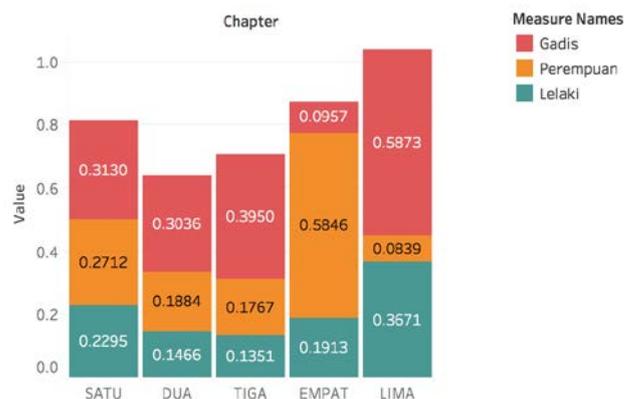


Fig. 2. A scaled proportion of usage frequency between *gadis*, *perempuan*, and *lelaki* throughout all chapters.

Based solely on the last two chapters, an early assumption can be made that *gadis* and *perempuan* might be used for different purposes or to construct different identities of female gender. Even though there is no unique pattern found for the male gender, it is still interesting to see how the three words interact with words or phrases around them and what patterns can be assumed from these interactions.

### B. Appraisal Theory

The notion of appraisal was first introduced into psychology by Magna Arnold [12], who defines it as the process through which the significance of a situation for an individual is determined. Appraisal brings about attraction or aversion, and emotion is associated with this "felt tendency toward anything intuitively appraised as good (beneficial), or away from anything intuitively appraised as bad (harmful)." [12]. ATTITUDE types in Appraisal Theory define the specifics of appraisal being expressed: AFFECT (personal emotional state), JUDGEMENT (social or ethical appraisal), and APPRECIATION (evaluation of phenomena).

APPRECIATION typically involves the evaluation of phenomena in relation to appearance or physical traits by reference to aesthetic principles and other systems of social

value. The attitudinal sub-system of JUDGEMENT encompasses meanings which serve to evaluate human behaviour positively and negatively by reference to a set of institutionalised norms which can be differentiated between personal judgements of admiration or criticism (SOCIAL ESTEEM) and moral judgements of praise or condemnation (SOCIAL SANCTION) [13], [14]. AFFECT is concerned with emotional response and disposition and is typically realised through mental processes of reaction and through attributive relationals of the types. Within this broad scope, the theory is more specifically concerned with the language of evaluation, attitude, and emotion, and with a set of resources which explicitly and interpersonally position a text's proposals and propositions.

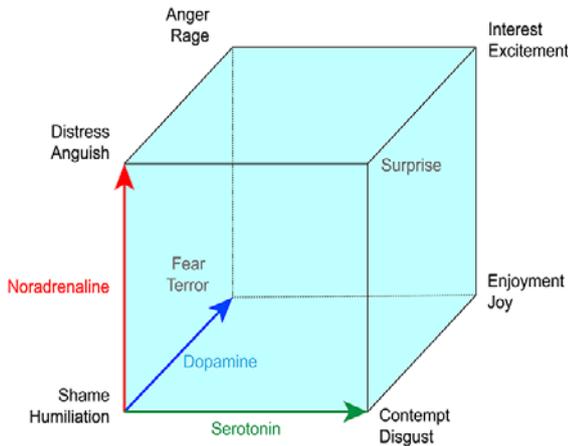


Fig. 3. Lövheim cube of emotion.

For this study, the eight AFFECT categories in the Lövheim cube of emotion in Fig. 3 [6] are together used as a dimensional model of emotion as a grounding base in building a contextual lexicon for attitude analysis. A combination of rule-based computational linguistic approaches and manual labelling are also used to support the categorisation and polarity of a word or phrase based on the sentence [5], [15] and to spot keywords based on common-sense knowledge [16].

### III. ANALYSIS

In analysing gender, certain words or phrases are usually used to convey femininity and masculinity, which are defined as gendered identities. Expanding from femininity and masculinity, the general characters of a gender are culturally construed and constructed, thus gender definitions are adaptable and malleable concepts. Because these characterisations are not innate, what represents male and female genders might vary socially, culturally, and generationally [17].

#### A. Concordances and Collocates

The main challenge of this study is to determine attitude-conveying terms and to continuously build a lexicon from the findings. The terms that construct gender identity description are often characterised by strings of adjectives and nouns [18]. Martin & White [14] also provided explanations and applications in which other word classes can be used to analyse appraisal of a certain target.

TABLE I: CONCORDANCE SAMPLES OF GADIS IN LELAKI HARIMAU

Post	Context (Pre)	Keyword	Context (Post)
2679	gambar di film layar tancap, atau disebabkan	gadis	<i>pujaan</i> [int+] hati berjalan dengan lelaki lain.
4357	mendidih, dan Anwar Sadat menjawab dagu	gadis-gadis	berseragam sekolah hingga mereka <i>merengek</i> [dis-]
28165	pacar. Tak ada yang lebih <i>memalukan</i> [hum-] bagi	gadis	di atas dua belas tahun tanpa dikenal siapa
44212	lampu penjual kacang rebus, ia <i>secantik</i> [app+]	gadis-gadis	dekorasi guci Cina, dengan rambut
44333	muda <i>perkasa</i> [app+] penakluk babi liar, serasa ia	gadis	<i>imut</i> [app+] <i>malu-malu</i> [hum-]. Orang-orang
54081	itu pun sanggup menghentikan <i>tangis</i> [dis-] si	gadis	, yang kini <i>menganga</i> [supø] dengan kata-kata

TABLE II: CONCORDANCE SAMPLES OF PEREMPUAN IN LELAKI HARIMAU

Post	Context (Pre)	Keyword	Context (Post)
14644	apa pun sebagaimana sering ia lakukan saat melihat	perempuan-perempuan	<i>cantik</i> [app+] dengan dandanan <i>provokatif</i> [prop-] macam begitu
38361	rasa dingin yang beku sekonyong mengentakkan si	perempuan	, yang tersadar dan <i>terlonjak</i> [supø]. Segera
14667	kini sedikit memejamkan matanya, diapit seorang	perempuan	<i>gembrot</i> [app-] yang datang bersama anak kecilnya,
19623	bicara dengan kompor dan panci sebagaimana biasa.	Perempuan	ini rada <i>sinting</i> [prop-], paling tidak begitulah pikir
43818	<i>bangsat</i> [prop-] mengetahui ketiadaan ayahnya, didengar Kasia membikin	perempuan	itu <i>menghardik</i> [cont-] untuk tidak membuat kata-kata
19728	<i>mengejutkan</i> [supø] itu bagi Mameh, menyadari	perempuan	ini bisa <i>bersedih</i> [dis-] pada suami yang sepanjang

TABLE III: CONCORDANCE SAMPLES OF LELAKI IN LELAKI HARIMAU

Post	Context (Pre)	Keyword	Context (Post)
47741	. Apa pula gunanya bunga-bunga untuk	lelaki	<i>bangsat</i> [prop-] <i>busuk</i> [ver-] ini, pikirnya. Namun
4844	mewariskan wajah <i>menarik</i> [app+] untuk mereka.	Lelaki	ini <i>tampam</i> [app+] selengkapnya, telah
37551	sana, dan tak ayal ini kerap bikin	lelaki	itu <i>cemburu</i> [rage-] buta, menghukum Nuraeni
16918	keluar dari sarang. Margio tak bakalan rela	lelaki	tua <i>celaka</i> [prop-] itu berbuat <i>baik</i> [prop-] untuk
17133	barangkali ia mewarisi harimau <i>keparat</i> [prop-] itu.	Lelaki	tua ini bertindak <i>bijak</i> [prop+] untuk tidak
5774	membiru, bahkan masih bisa <i>bersedih</i> [dis-] pada	lelaki	yang jelas <i>tak setia</i> [ver-] itu. Dan orang

Neviarouskaya, Prendinger and Ishizuka [19] developed an ATTITUDE lexicon database for ATTITUDE analysis which contains sentiment-conveying adjectives adverbs, nouns, and verbs annotated by sentiment polarity, polarity scores, and weights that was manually extended using attitude labels. The study proposes methods to automatically generate and score a new sentiment lexicon, called SentiFul, and expand it through direct synonymy relations and

morphologic modifications with known lexical units. However, this is not applicable for the Indonesian language since most of the lexicons that convey sentiment don't have evident modification patterns. Follow-up studies will focus specifically on finding an effective and scalable method to detect sentiment-conveying terms in Indonesian.

Tables I, II, and III show the concordance samples for each word with ATTITUDE labelling based on results from collocations. Here, words and phrases are marked with attitudinal expressions of 'app' for APPRECIATION; 'nor', 'cap', and 'ten' for NORMALITY, CAPACITY, TENACITY respectively in the SOCIAL ESTEEM group; 'ver' and 'prop' for VERACITY and PROPRIETY respectively in the SOCIAL SANCTION group; all of JUDGEMENT [14]; and eight AFFECT categories [6]. '+', '-', and 'ø' indicate positive, negative, and undecided polarity, respectively.

B. Attitude Analysis (Appraisal Theory)

1) Appreciation analysis

A striking disparity is illustrated in Table IV in terms of variation and frequency of attitude-conveying terms between female gender (*gadis* and *perempuan*) and male gender (*lelaki*) with a single occurrence token. No negative sentiment term was found interacting with either *gadis* or *lelaki* and there is no interaction between the female gender nodes and male gender node as presented in Fig. 4. The word *cantik* (beautiful) shares interaction with both *gadis* (thicker line width indicates more in frequency) and *perempuan*.

TABLE IV: APPRECIATION ANALYSIS TABLE (CATEGORY AND POLARITY)

gadis		
Word	Occurrence	Polarity
<i>Cantik</i>	11	positive
<i>Indah</i>	1	positive
<i>menarik</i>	1	positive
<i>Imut</i>	1	positive
<i>Manis</i>	1	positive
perempuan		
Word	Occurrence	Polarity
<i>Cantik</i>	7	positive
<i>gembrot</i>	1	negative
<i>semrawut</i>	1	negative
lelaki		
Word	Occurrence	Polarity
<i>tampan</i>	1	positive

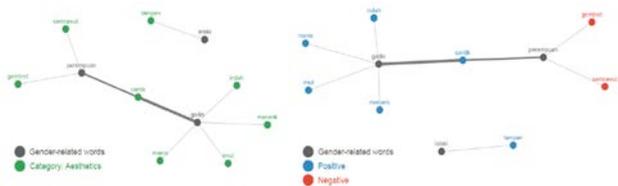


Fig. 4. Network graph of APPRECIATION analysis (category and polarity).

2) Judgement analysis

It can be seen in Tables V to VII that there is no PROPRIETY praise or condemnation associated with *gadis*, whereas interactions with *perempuan* and especially *lelaki* are quite dominated by it.

Both *perempuan* and *lelaki* are surrounded mainly by negative PROPRIETY words in Tables VI and VII. Provided that this is in line with the novel's theme, it is interesting to

see the contrast between *perempuan* and *gadis* specifically in this aspect.

TABLE V: JUDGEMENT ANALYSIS TABLE OF GADIS (CATEGORY AND POLARITY)

Category	gadis		
	Word	Occurrence	Polarity
tenacity	<i>cengeng</i>	1	negative
	<i>gesit</i>	1	positive
	<i>cekat</i>	1	positive
	<i>keberanian</i>	1	positive
capacity	<i>cerdas</i>	1	positive
	<i>ingusan</i>	1	negative
	<i>(hilang) kewarasan</i>	1	negative
normality	<i>riang</i>	2	positive

TABLE VI: JUDGEMENT ANALYSIS TABLE OF PEREMPUAN (CATEGORY AND POLARITY)

Category	perempuan		
	Word	Occurrence	Polarity
propriety	<i>baik-baik</i>	1	positive
	<i>binal</i>	1	negative
	<i>begundal</i>	1	negative
	<i>sundal</i>	1	negative
	<i>genit</i>	1	negative
tenacity	<i>cengeng</i>	1	negative
	<i>pemalas</i>	1	negative
	<i>bernyali</i>	1	positive
	<i>sinting</i>	3	negative
veracity	<i>provokatif</i>	1	negative

TABLE VII: JUDGEMENT ANALYSIS TABLE OF LELAKI (CATEGORY AND POLARITY)

Category	lelaki		
	Word	Occurrence	Polarity
propriety	<i>baik</i>	1	positive
	<i>bijak</i>	1	positive
	<i>keparat</i>	1	negative
	<i>bangsat</i>	1	negative
	<i>bengis</i>	1	negative
	<i>jahil</i>	1	negative
	<i>mata-keranjang</i>	1	negative
	<i>keji</i>	1	negative
	<i>tak kenal adat</i>	1	negative
	<i>celaka</i>	1	negative
tenacity	<i>gagah berani</i>	1	positive
	<i>bertelete-tele</i>	1	negative
capacity	<i>cerdas</i>	1	positive
	<i>cermat</i>	1	positive
veracity	<i>busuk</i>	1	negative
	<i>pendusta</i>	1	negative
	<i>tak setia</i>	1	negative

In Fig. 5, only the word *cengeng* (whiny) is shared between *gadis* and *perempuan*, while the word *cerdas* (smart) is shared between *gadis* and *lelaki*.

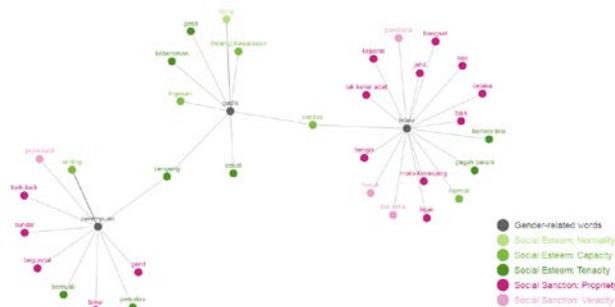


Fig. 5. Network graph of JUDGEMENT analysis (category).

In terms of sentiment, in Fig. 6, *gadis* displays interaction with an equal ratio, but it is lopsided for both *perempuan*

and *lelaki* with a tendency towards the negative side. It can be inferred from JUDGEMENT analysis that both *perempuan* and *lelaki* pose a similar inclination even though the lexicons used for each word are different. For instance, the word *binal* (wild) is commonly used to describe female characteristics, and is rarely found to describe male characteristics. There is also a clear contrast in inclination between *gadis* and *perempuan* or *lelaki*, particularly in the PROPRIETY aspect.

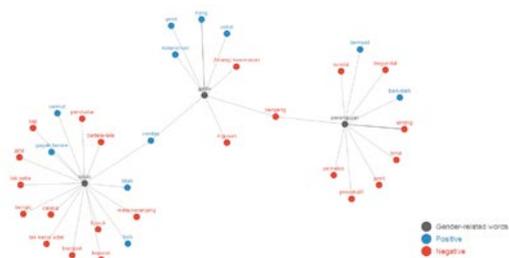


Fig. 6. Network graph of JUDGEMENT analysis (polarity).

### 3) Affect analysis

This study uses the Lövheim [6] cube of emotion to construct eight categories of AFFECT. There are many tokens indicating the semantic theme of distress or anguish for *gadis* and *perempuan* in Tables VIII and IX. There are tokens found for all eight categories proposed by Lövheim surrounding *gadis* in Table VIII, as well as two additional tokens, *kosong* (empty/numb), *tak hangat* (no warmth/dull), which don't quite fit any of the emotions proposed by Lövheim. Both tokens in English might describe something close to distress or anguish, but in Indonesian they are used to describe having no emotion at all or a completely bland feeling. Arguably, these characteristics evoke different emotions than do distress or anguish; therefore, they should be categorised differently.

TABLE VIII: AFFECT ANALYSIS TABLE OF GADIS (CATEGORY AND POLARITY)

Category	Polarity	<i>gadis</i>	
		Word	Occurrence
distress/anguish	negative	<i>merengek</i>	1
		<i>tangis (bersimbah air mata)</i>	2
		<i>(rasa) sakit hati</i>	3
		<i>patah hati</i>	2
		<i>berserah</i>	1
interest/excitement	positive	<i>pujaan</i>	1
		<i>terpuja</i>	1
		<i>tergila-gila</i>	2
enjoyment	positive	<i>cinta</i>	6
		<i>lapang</i>	1
		<i>senang</i>	1
anger/rage	positive	<i>(tak) sabar</i>	1
		<i>dongkol</i>	1
anger/rage	negative	<i>penolakan</i>	1
		<i>memalukan</i>	1
		<i>malu-malu</i>	1
contempt/disgust	negative	<i>berlumut</i>	1
surprise	undecided	<i>menjerit</i>	1
		<i>menganga</i>	1
numbness	negative	<i>kosong</i>	1
		<i>(tak) hangat</i>	1
fear/terror	negative	<i>waswas</i>	1

There is also an opposing pattern between *gadis* and *lelaki* in terms of interactions with judgement-conveying terms in Tables V and VII and emotion-conveying terms in Tables VIII and X, while *perempuan*'s interactions with the terms in both graphs are quite similar in Tables VI and IX. The word *gadis* has the most interactions with emotion-conveying terms, but the least with judgement-conveying terms, and vice-versa for *lelaki*.

The words *nafsu/berahi* (lust/passion) found interacting with *perempuan* and *lelaki* in Tables IX and X relate with excitement; however, the context might vary greatly for this word. Similar to *nafsu/berahi* (lust/passion), the words *terlonjak* (jumped), *menganga* (gawk), and *menjerit* (scream) can vary on the sentiment scale greatly depending on its pragmatic context. Therefore, an 'undecided' label is assigned for their polarity as displayed in Fig. 8. This label proposes further study in the Appraisal Theory weighting of both categorisation and polarity to provide a more accurate sentiment and theme analysis.

TABLE IX: AFFECT ANALYSIS TABLE OF PEREMPUAN (CATEGORY AND POLARITY)

Category	Polarity	<i>perempuan</i>	
		Word	Occurrence
distress/anguish	negative	<i>duka</i>	3
		<i>senggukan</i>	1
		<i>isak</i>	1
		<i>sedih</i>	3
		<i>menderitakan</i>	1
		<i>menangis</i>	1
		<i>membuncah</i>	1
interest/excitement	positive	<i>suka</i>	1
		<i>memuja</i>	1
		<i>berahi*</i>	1
enjoyment	positive	<i>tertawa</i>	2
		<i>senyum</i>	1
		<i>terkekeh-kekeh</i>	1
anger/rage	negative	<i>memaki</i>	1
		<i>kemarahan</i>	1
contempt/disgust	negative	<i>menghardik</i>	1
		<i>membosankan</i>	1
surprise	undecided	<i>menjerit</i>	1
		<i>terlonjak</i>	1

TABLE X: AFFECT ANALYSIS TABLE OF LELAKI (CATEGORY AND POLARITY)

Category	Polarity	<i>perempuan</i>	
		Word	Occurrence
distress/anguish	negative	<i>payah</i>	1
interest/excitement	positive	<i>terpana</i>	1
		<i>nafsu*</i>	1
enjoyment	positive	<i>kehangatan</i>	1
		<i>senyum</i>	1
anger/rage	negative	<i>ganas</i>	1
		<i>cemburu</i>	1
shame/humiliation	negative	<i>penolakan</i>	1
		<i>canggung</i>	1
contempt/disgust	negative	<i>menyebalkan</i>	1

The striking difference between female and male gender representation in the AFFECT analysis is the number of words as seen in Table X. Even though this might be due to keeping in line with the theme of Kurniawan's novel, it poses a question of whether this pre-existing stereotypical pattern that the female is considered to be more emotional than the male can also be found across novel genres or even

across literature in general.

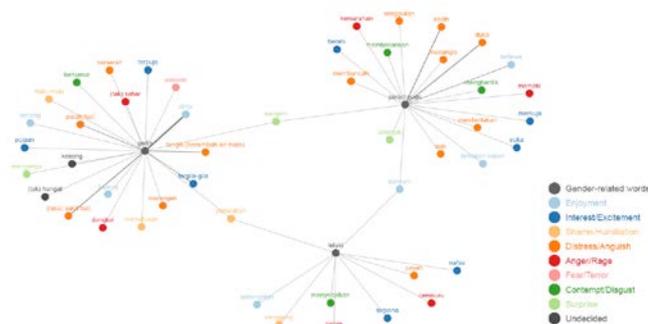


Fig. 7. Network graph of AFFECT analysis (category).

A few words are shared between the keywords, as seen in Fig. 7 and 8. The word *penolakan* (rejection) is shared between *gadis* and *lelaki*, while *menjerit* (scream) is shared between *gadis* and *perempuan*, and *senyum* (smile) is shared between *perempuan* and *lelaki*. In terms of the statement in Fig. 8, there is no singular significant pattern found.

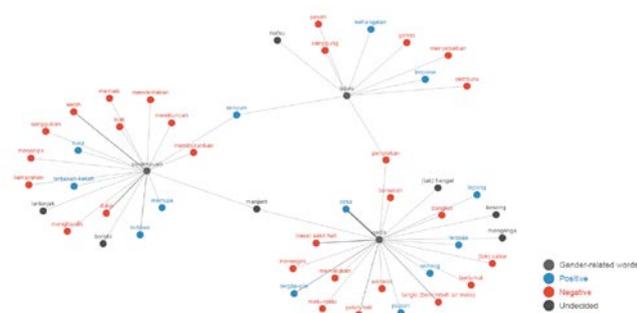


Fig. 8. Network graph of AFFECT analysis (polarity).

#### IV. CONCLUSION AND FUTURE WORK

The combination of Appraisal Theory and computational methods employed in this study generate typical and atypical interactions by comparing the syntagmatic and paradigmatic structures, semantics, concordances, and collocates. Eka Kurniawan's style of using vivid and descriptive keywords is an interesting aspect to study in seeking to understand patterns underlying gender representation in literature. Pre-existing stereotypical descriptions of female and male genders are common in the data, such as that of appearance, which are associated more with female gender words and emotion-conveying terms that interact less frequently with male gender words.

The existence of *gadis* and *perempuan* as female gender-related words displays interesting characterisations within a gender. For instance, it can be implied from the data that associating praise and condemnation with *gadis* is uncommon, while it can be found in many interactions with *perempuan*. Particularly for the male gender, the sentiment found in its interaction with judgement-conveying terms is distinct in condemnation. Even though there is not much of a disparity between *gadis* and *perempuan* in the AFFECT analysis, it would still be interesting to further investigate whether there is a distinguishing theme or style of speech associated with the two words.

The contrasting usage frequency between *gadis* and *perempuan* in the last two chapters indicates the different

purpose or characterisation by using one or the other. The word *gadis* usually only differs in usage with *perempuan* because it describes females who are not yet married. However, the more specific theme difference surrounding the two words has never been discussed with the support of evidence from quantitative analyses. To conduct a thorough and general gender analysis in the literature, every gender-related aspect need to be considered and analysed. This study also shows how the scoring and weighting system of Appraisal Theory elements for both categorisation and polarity is paramount, because each word imposes complex meanings that might overlap in category or polarity to different degrees. Further analysis might provide not just potential instances or patterns generated from word interactions, but also overall accurate identity categorisation and profiling.

Early experiments to quantify ATTITUDE analysis focusing on distinct ATTITUDE types were conducted by Taboada and Grieve [20]. Taboada and Grieve determined a potential value of adjectives for APPRECIATION, JUDGEMENT, and AFFECT by calculating the PMI with the pronoun-copular pairs such as '*I was (affect)*', '*he/she was (judgement)*', and '*it was (appreciation)*'. Whitelaw *et al.* [21] also used a machine learning approach with fine-grained semantic distinctions in attitude types and orientation, utilising a 'bag of words' to classify movie reviews. However, these endeavours are mainly designed for English language usage. A robust and comprehensive dictionary in Indonesian with word classes, POS rules, and scoring and weighting system will be a great contribution to support contextual and more accurate research in linguistics, with the aim to provide natural language understanding for machine learning and artificial intelligence purposes' contextual analysis of language in Indonesia.

#### REFERENCES

- [1] C. David, *An Encyclopedic Dictionary of Language and Languages*, Oxford, UK: Blackwell. Print, 1992.
- [2] J. Howard, *Key Terms in Linguistics*, London: Continuum. Print, 2007.
- [3] S. J. McHardy, *Trust the Text: Language, Corpus and Discourse*. London: Routledge. Print, 2005.
- [4] M. A. K. Halliday, *An introduction to Functional Grammar*, London: Edward Arnold, 1994.
- [5] M. Karo and S. Pulman, "Sentiment composition," in *Proc. Natural Language Processing International Conference on the Recent Advances*, pp. 378-382, 2007.
- [6] L. Hugo, "A new three-dimensional model for emotions and monoamine neurotransmitters," *Medical Hypotheses*, vol. 78, no. 2, pp. 341-48, 2011.
- [7] K. Eka, *Lelaki Harimau*. Jakarta: Gramedia Pustaka Utama, 2004.
- [8] R. C. Team. (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. [Online]. Available: <https://www.R-project.org/>
- [9] M. Chris and H. Schütze, *Foundations of Statistical Natural Language Processing*, Cambridge, MA: MIT Press, 1999.
- [10] B. Kenneth and P. Nulty. (2017). Quanteda: Quantitative analysis of textual data. R package version: 0.9.9-24. [Online]. Available: <https://github.com/kbenoit/quanteda>
- [11] T. Kamada and S. Kawai, "An algorithm for drawing general undirected graphs," *Information Processing Letters*, 1989.
- [12] A. Magda, *Emotion and Personality*, New York: Columbia University Press, 1960.
- [13] I. Rick, S. Feez, and P. R. R. White, *Media Literacy, Sydney, Disadvantaged Schools Program*, NSW Department of School Education, 1994.
- [14] J. R. Martin and P. R. R. White, *The Language of Evaluation: Appraisal in English*, Palgrave, London, UK, 2005.

- [15] V. S. Subrahmanian and D. Reforgiato, "AVA: Adjective-verb-adverb combinations for sentiment analysis," *Intelligent Systems*, IEEE, vol. 23, no. 4, pp. 43-50, 2008.
- [16] H. Liu, H. Lieberman, and T. Selker, "A model of textual affect sensing using real-world knowledge," in *Proc. IUI-2003*, pp. 125-132, 2003.
- [17] P. Baker, "No Effeminate Please: A Corpus-based analysis of masculinity via personal adverts in gay news/times 1973-2000," in *Masculinity and Men's Lifestyle Magazines*, B. Benwell, Ed. Blackwell. Oxford, 2003.
- [18] P. Bruthiaux, *Me Tarzan, You Jane: Linguistic Simplification in "Personal Ads" Register*, In Biber and Finegan (1994a), 136-54, 1994.
- [19] A. Neviarouskaya, P. Helmut, and M. Ishizuka, "SentiFul: Generating a reliable lexicon for sentiment analysis," in *Proc. the International Conference on Affective Computing and Intelligent Interaction*, IEEE, Amsterdam, Netherlands, 2009, pp. 363-368.
- [20] M. Taboada and J. Grieve, "Analyzing appraisal automatically," in *Proc. AAAI Spring Symposium on Exploring Attitude and Affect in Text*, pp.158-161, 2004.
- [21] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in *Proc. the 14th ACM International Conference on Information and Knowledge Management*, CIKM, Bremen, Germany, pp. 625-631, 2005.



**Lalitia Apsari** was born in Jakarta, Indonesia. She received her bachelor of architecture (B.Arch) degree in 2006 from University of Indonesia with an undergraduate thesis on philosophy and semiotic studies on Prambanan Temple's iconography using phenomenology as the main analysis approach.

She attended the Digital Humanities at Oxford Summer School (DHOxSS) focusing on Data Science for the Humanities. She also participated in a wide range of courses such as "Visualizing Japan (1850s-1930s): Westernization, Protest, Modernity", a collaboration between HarvardX and MITx, and "The science of everyday thinking" by University of Queensland. She has been a co-curator for TEDxJakarta since 2010, an advocate of digital humanities movement in Indonesia, and a member of data science Indonesia community. Her multidisciplinary research interests include phenomenology and deconstruction, cognition and appraisal theory, the computation of meaning from corpora, and digital humanities.

Mrs. Lalitia Apsari is currently pursuing her graduate study in Linguistics at University of Indonesia while being an assistant researcher to develop comprehensive and contextual language resources for humanities studies in Indonesia. She is also completing her research on the manifestation and change of human cognition as reflected in a large text corpora of newspaper articles in Indonesia collected since 2004.