Language use variation of L2 writers in weblog across different gender and genres

Dian Puspita[1], Suprayogi[2] ✉

[1, 2]English Literature, Universitas Teknokrat Indonesia, Indonesia

Article Info	Abstract
	Despite the growing interest in investigating learners' corpora, surprisingly little research has been conducted on the language use of L2 writers and its relation to the gender and genres in writing. Therefore, this study was aimed to find out the variation of language use in different genres or gender in weblogs, one of popular modes of computer-mediated communication (CMC). The study was done by conducting multivariate analysis using R program to weblog entries from a sample balanced of author gender (female or male) and weblog genre (diary or filter). Taking linguistic preferential features by Argamon et al (2003) and Pennebaker (2011) as dependent variables, the effect of genres or gender toward the use of the features was analyzed. The results showed that significant effects of several features can be considered as predictors. Personal pronouns and hedges (I think, and I believe) were found as predictors for diary, while the indefinite articles a/an and numbers were found as predictors for filter. As for the different language use by gender, female predictors were personal pronoun, verbs, negation, certainty words, and hedges. Meanwhile, the indefinite articles a/an, numbers, and preposition were the predictors of male writers.

INTRODUCTION

Technology has emerged a new approach of communication called computer-mediated communication (CMC). CMC is a way of communication that takes place between human beings via the instrumentality of computers (Herring & Paolillo, 2006). CMC broadens its audience from all around the world, and it provides language with a whole range of new spaces in which to work and play (Pemberton & Shurville, 2000). Undoubtedly, the new means of communication resulted in new and rich characteristics of languages of varied users (Bodomo, 2010; Ess & Sudweeks, n.d.; Herring, 2002). For its broad users and the easy access to

manifest the data, CMC has been seen as linguistic corpus data and seen as ethnographic observation of naturally occurring interactions (Herring, 2002). This ethnographic observation is employed as an approach in sociolinguistic research, especially online ethnography as one sociolinguistic issue in CMC (Androutsopoulos, 2006; Suprayogi, 2019)

One of sociolinguistic issues which is cherished and interesting to study from CMC is language and gender. A number of studies paying a considerable attention to the effect of gender stereotypes toward linguistic behavior, especially in text-based CMC, have been done by many scholars (Argamon et al., 2003; Herring & Paolillo, 2006; Koppel et al., 2004; Pennebaker,

2011; Samar & Shiazizadeh, 2010; Thomson & Murachver, 2001). These studies found the patterns of certain linguistic behavior toward each gender, so called gender preferential linguistic features (Koppel et al., 2004; Pennebaker, 2011) or predictor of author gender (Herring & Paolillo, 2006; Koppel et al., 2004; Samar & Shiazizadeh, 2010). These studies agreed to the gender influence toward language use and are in line with gender stereotype which were affirmatively found in gender's topic association (e.g., Argamon et al., 2003; Colley & Todd, 2002; Janssen & Murachver, 2004; Thomson & Murachver, 2001). As the stereotype believed for each gender's role, female speak and hear a language of connection and intimacy, while male speak language and hear a language of status and independence (Tannen as cited in Mesthrie et al., 2009)

Evidently, these stereotyped gender' roles were also present in topics performed by each gender in written language (Colley & Todd, 2002; Mooney & Evans, 2015; Thomson & Murachver, 2001). Females tend to present topics related to personal and emotional things which are more involving and interacting with the audience. Males, on the other hand, presented factual and informative topics in their writing. Herring & Paolillo (2006) found that the use of personal pronouns, as in gender preferential linguistic features, was not relevant. It was explained that the use of certain pronouns was not determined by the gender, but by the topics writers chose. Furthermore, it was also shown that females did not produce the writings related to personal and emotional topics only, but they were found to write factual and informative topics also. It occurred to males vice versa. Similarly, gender preferential linguistic element was used to investigate the research papers written by non-native speakers of English (Samar & Shiazizadeh, 2010). They found that the difference between males and females in the frequency of using those features was not statistically significant. This non-significant difference shows that either the confinements of genre or those of using a second language or both are keeping 12 writers from expressing their gender to its fullest capacity in the texts they produce.

From the previous studies mentioned, it can be seen that there were two groups proving different effects of genres and gender toward language use; one revealed that gender brings the more significant effect on language use, while the other found that it was not gender effect, but more genre effect. Therefore, as previous studies reviewed above arrive at inconclusive results, it is of interest to compare whether gender or genre affect more in non-native English writing in weblog entries using the linguistic features by Argamon et al., (2003), and also linguistic features by (Pennebaker, 2011). This study is addressed to answer the following research questions: 1) Do male blog authors write certain English language use differently from female blogs authors in different text type (genre)? 2) To what extent do gender and genre affect language use?

METHODS

The corpus from 68 English-written blogs by Indonesian writers was taken from the list of blog rank in Indonesia which was based on Google page rank, number of subscribers, back link, back tweets, social bookmarks, engagement scores and the traffic rank. To reduce the data based on the gender, single-authored blogs were selected. It turned out that not all of those weblogs were accessible and written in English. As a result, the written collection was only collected as many as 77 entries including two genres for both genders. Therefore, to reach the data needed, snowball sampling was done. At the end, a number of other web-blogs obtained from the link connected from the previous web-blogs and the entries were 120 entries. This corpus based study was currently significant in linguistic studies (Puspita & Pranoto, 2021; Puspita, 2019a, 2019c; Sari & Gulö, 2019).

Table 1 Weblog data sample

Gender	Text type	Number of entries
Male	Filter	30
Male	Diary	30
Female	Filter	30
Female	Diary	30

The frequency counting of each linguistic feature and the total number of words for each entry was done automatically by using Antconc 3.5.7 software. Adapting the research framework by (Herring & Paolillo, 2006), this study used the linguistic features by Argamon et al., (2003)

Table 2 Gender Preferential Linguistic Features by Argomon, Kopel, and Simon

Female preferential features	Male preferential features
First-person singular (I,me,my,mine)	The determiners (the and a/an)
First-person plural (we, us, our, ours, and 's in let's)	Demonstratives (this, these, that, those)
Third-person singular (forms of she and he)	numbers (1, 2, 1000, one, two, thousand, first, second, etc.)
Third-person plural (they, them, their, theirs)	other quantifiers
	the possessive pronoun its

The later gender preferential linguistic features found by (Pennebaker, 2011) were used as well to be compared. Differently, not all features were taken because in the previous study, Pennebaker built the corpus not only from web-blog but also other sample texts such as play scripts, and books which were moderately irrelevant to web-blog genres. Therefore, some features were not taken into account.

Table 3 Gender Preferential Linguistic Features by Pennebaker

Female preferential features	Male preferential features
Personal pronouns	Big words
Verbs (including auxiliary verbs)	Nouns
Negative emotion (especially anxiety)	Prepositions
Negations (no, not, never)	Numbers
Certainty words (always, absolutely)	Words per sentence
Hedge phrases ("I think," "I believe")	Swear words

After counting the frequencies, to explore the relationship between genre and gender toward language use in weblog, a multivariate analysis of the features in the weblog entries was done. The writer adopted this quantitative approach in order to evaluate the applicability of the quantitative claims made in Argamon and Koppel's work and also Pennebaker's work. To be able to do so, first dependent and independent variables should be set.

RESULTS AND DISCUSSION

The results showing the relationship between genre and gender toward language use in weblog from logistic regression analysis using R are presented below. Two logistic regressions were used to analyze the effect of genre and gender toward the female-preferential features and male-preferential features.

Female Preferential Features

Argamon et al's Preferential Features

The result of logistic regression shows that all of the feature-genre interactions were significant (p-value ≤ 0.05), indicating that all of the features put different rates of use in the two genres. In other words, all of the pronoun features show a different distribution by genre.

The output of the statistical model that corresponds to these coefficients observations is presented in table 4. The four columns of this table are: Parameter, giving the name of each effect in the model; its Estimate on the logit scale; the p value associated with that parameter; and a significance code (*p < 0.05, **p < 0.01, ***p < 0.001).

Table 4 Logistic regression model for Argamon et al's (2003) female-preferential features

Parameter	Estimate	p	Significance
gnr1: feat1	0.6507058	0.0001212072	***
gnr1: feat2	1.4193402	0.0007055561	**
gnr1: feat3	1.8821095	0.0000000000	***
gnr1: feat4	2.6443732	0.0000000000	***
We: gnd1	0.3882654	0.004291693	**
She: gnd1	0.9569321	0.011325981	*
He: gnd1	0.1158700	0.643799557	-
They: gnd1	0.3882869	0.024013417	*

Note: 'gnr1' = diary; 'gnd1' = female

To better explain the statistical model, Figure 1 presented the observed frequencies of each of the features. Of all features, first-person singular forms were the most frequent, followed by third-person plural, first-person plural, third-person male and third-person female respectively. Diary entries showed much greater use of first-person singular than filter entries did.
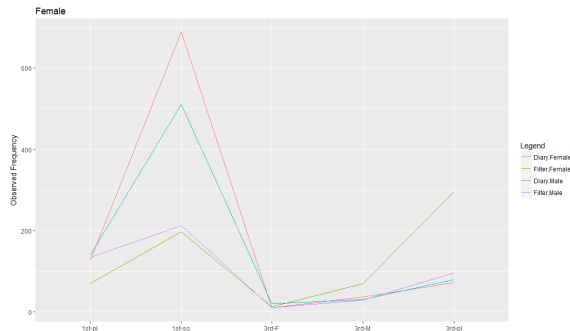

Figure 1: Observed frequencies of Argamon et al's (2003) female-preferential features

Pennebaker's Female-Preferential Features

Logistic regression result showed that for the feature-genre interactions, hedge phrases and personal pronouns were significant (p-value ≤ 0.05). Meanwhile, negation and verbs were not significant because the p-values were not less than 0.05 (0.2190986 and 0.5321200 respectively). For the feature-gender interaction, all features were significant. For the interaction between genre and gender, only negations were significant. Other features were not significant because the p-values were not less than 0.05. The output of the statistical model that corresponded to these coefficients observations is presented in Table 5.

Table 5 Logistic regression model for Pennebaker's female-preferential features.

Parameter	Estimate	*P*	Significance
gnr1: feat1	-0.58812431	-0.9317426	***
gnr1: feat2	0.09253381	0.2190986	-
gnr1: feat3	-0.58298179	-1.4395025	***
gnr1: feat4	0.21488852	0.5321200	-
hegde phrases: gnd1	-1.3656375	-1.7312069	***
negations: gnd1	-0.1208386	-0.2731208	***
personal	-0.2345114	-0.5554662	***

pronouns: gnd1			
verbs: gnd1	-0.1026050	-0.2434741	***

Note: 'gnr1' = diary; 'gnd1' = female

To better explain the statistical model, Figure 2 presents the observed frequencies of each of the features. Of all features, verbs were the most frequent, followed by personal pronoun, negations, hedges, and certainty words respectively. Diary entries showed much greater use of personal pronoun and verbs than filters did.
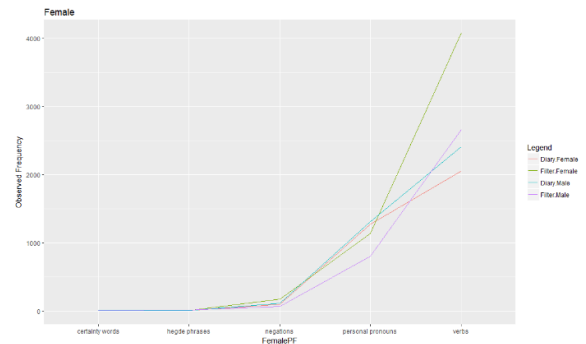

Figure 2: Observed frequencies of Pennebaker's female-preferential features

Male Preferential Features

Male-Preferential Features Argamon et al's Male-Preferential Features

The logistic regression result showed that only two feature-genre interactions were significant; a/an and numbers. These two interactions were also present exactly the same to the feature-gender interaction. Only a/an and numbers showed significant interactions. In other words, only half of the interaction parameters in this model were significant, meaning that the overall differences in the gender and genre distribution of the features were not so great. The output of the statistical model that corresponds to these coefficients observations was presented in Table 6.

Table 6 Logistic regression model for Argamon et al's male-preferential features

Parameter	Estimate	*p*	Significance
gnr1: feat1	0.7214501	3.032241e-12	***
gnr1: feat2	0.1549926	1.732656e-01	-

gnr1: feat3	0.3801223	3.561150e-01	-
gnr1: feat4	0.2783635	1.492538e-02	*
A/An: gnd1	0.36959581	3.032241e-12	***
Dem: gnd1	0.08278789	5.089956e-01	-
Its: gnd1	0.44848079	3.076981e-01	-
Number: gnd1	0.57415524	1.795531e-05	***

Note: 'gnr1' = diary; 'gnd1' = female

From the logistic regression, it can be concluded that the features a/an, and numbers were significant for both genre and gender. Figure 3 showed the observed frequencies of use of each of the male-preferential features. With regard to genre, there was all overall positive main effect associating filter genre with the hypothesized male-preferential features. However, only a/an and numbers were significant. Because the was taken by the statistical program as the reference category, Table 4.6 did not include a significance measure for it, but a significant positive correlation with the filter genre can be inferred from the main effect for genre and the observed frequencies of the in Figure 3.
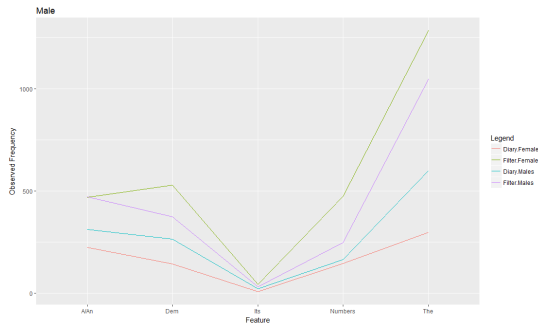

Figure 3: Observed frequencies of Argamon et al's male-preferential features

Pennebaker's Male-Preferential Features

The logistic regression result showed that for the feature-genre interactions, number was significant. Meanwhile, preposition was not significant. On the contrary, for the feature-gender interaction, preposition showed significant interaction, but number was not. The output of the statistical model that corresponds to these coefficients observations was presented in Table 7.

Table 7 Logistic regression model for Pennebaker's male-preferential features.

Parameter	Estimate	p	Significance
gnr1: feat1	-0.0012685	-0.0131970	***
gnr1: feat2	0.00919091	0.8360077	-
Number: gnd1	-0.1244647	2.745531e-01	-
Preposition: gnd1	0.1942836	9.638487e-05	***

Note: 'gnr1' = diary; 'gnd1' = female

To better explain the statistical model, Figure 4 presents the observed frequencies of each of the features. Of all features, noun was the most frequent, followed by preposition and numbers. Filter entries showed much greater use of all features than diary entries did.
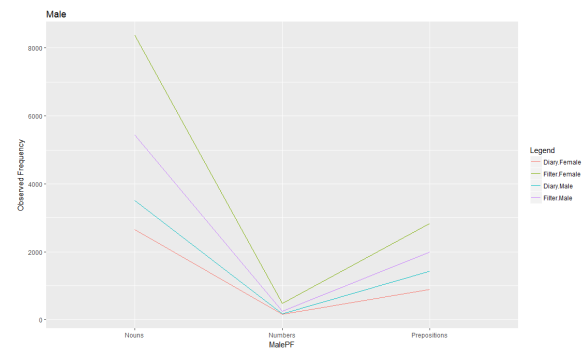

Figure 4: Observed frequencies of Pennebaker's male-preferential features

Discussion

The results from the logistic regression showed that there was different English language use performed by both the diary and filter authors. Also, different preferences were performed by female and male authors.

Based on Argamon et al., (2003) and Pennebaker (2011) preferential features, the findings showed that diary authors use great number of pronouns. It was due to the characteristic of diary genre itself. As its definition and classification, diary covers the text type with topics on the basis of a blog's overall purpose such as to report and comment on the author's own life Herring & Paolillo (2006). In other words, diary basically tells about people. Therefore, it is unsurprising that the great use of pronoun was found in diary weblog genre. Following are the

examples of the diary entries title from the data sources.

Table 8 Diary entries

Genre	Gender	Title
Diary	Female	Back On Blogging
Diary	Female	Ex And The City
Diary	Female	Life, So Far
Diary	Female	Paris Day 1
Diary	Male	First Semester
Diary	Male	Here In My Car
Diary	Male	21 Candles
Diary	Male	Semi Late New Year Resolution

Different from diary, filter authors were found to use a/an, dem, its, and number. It is also due to the characteristic of filter genre itself. Filter genre is defined and classified as covering the text types with topics containing information or events external to the author (Herring & Paolillo, 2006). In other words, it tells about thing outside the writer's personal worlds, such as technology, politic, science, etc. The following titles were the example of the filter entries containing those topics.

Table 9 Filter entries

Genre	Gender	Title
Filter	Female	Logic Of Religious Violence
Filter	Female	Idiopathic
Filter	Female	Screen Policy In South Korea
Filter	Female	World Retail Congress Asia Pacific 2013
Filter	Male	Diplomatic Dipsomania Two Sides Of IDX Head Tito Sulistio In Balcony Collapse
Filter	Male	Political Update
Filter	Male	Implementing Lua Scripting And Visual Scripting For Game Engine
Filter	Male	Maritime History Part 2

Writing about such topics outside writers' personal world means that they used a great number of nouns in the entries as it was shown in figure 4. Therefore, the finding about the great use of aforementioned preferences; a/an, dem, its, and number was also relevant. It is because these features were grammatically positioned before nouns.

As for different preferences in gender, it was found that Argamon et al. (2003) female-preferential features were accurate. The result showed that female authors used more pronoun than male authors did. All pronoun preferences by Argamon et al. (2003) were found significant except for third-person singular male. The sociolinguistic factor of Indonesian language may influence this finding. In Indonesia, the gender marker was present in phonological and morphological levels (Triyono, 2003). However, the gender marker in morpheme was not present in the 3rd-person singular; there was no difference in female and male 3rd-person singular in Indonesian language. Moreover, just a few female writers, in diary genre, wrote about their personal topics which were related to male 3rd-person. In fact, they wrote more about their personal life.

For all personal pronouns by Pennebaker (2011), it was also found that female use more pronouns more significantly than male. Argamon et al. (2003) argued that the use of personal pronoun in telling stories was a strategy to encode the relationship between the writer and the reader. In this study, this strategy was also used by female to engage the readers to the stories they were telling. It proved gender stereotype which believes that females like to write about topics related to personal matters and also about building connections with the readers (Domínguez-Rué, 2012). Female authors were also found to use negation, and hedge phrases more than male-authors did. This finding proved that gender stereotype for female existed in this study. The stereotype believed that females are more tentative and uncertain in expressing their ideas (Mesthrie et al., 2009), so they use certain expression such as negation and hedges. Here are two examples of hedges used by female author.

. ... I think was too long and too far wandering about life outside my blog.
(Diary - Leaping Year)
..... It 's going to be a very productive day , I think.
(Diary - Mapple)

The examples above showed the function of hedges as Pennebaker (2011) called "acknowledging". Instead of expressing in confidence their ideas; it was too long; and it's going to be a very productive day; the writers use "I think" to implicitly acknowledge that there are different views on this, and you may indeed come to a different conclusion, but my own personal belief is this one.

For male authors, the result showed that they use a/an, number and preposition more that female authors did. It was followed by the use of noun by male authors. It also proved the gender stereotype which put male as gender that had a tendency to prefer more generic pronouns rather than personal pronoun. It is because male are believed to talk and write about things instead of about people (Tillery, 2005). Even though the data taken was from two genres and males wrote diary entries, it was found out that they used lack of personal pronoun compared to female writers. Instead, they preferred using generic pronouns.

CONCLUSION

Based on the study conducted, the findings could answer the two research questions posed in this study. The first finding was that male blog authors write differently from female blogs authors. To see the difference, the findings shows to what extent genre and gender affect language use which was answering the second research question.

For weblog genre, based on Argamon et al., (2003) female-preferential features, diary genre uses more pronoun (first-person singular, first person plural, third-person singular, and third-person plural) than filter genre does. Moreover, all feature-genre interactions are significant, meaning that the features are found as genre predictor for diary. Also, based on Argamon et al. (2003) male-preferential features, it is found that filter genre uses a/an, dem, its, and number more than diary genre. However, there are only two significant p-values, a/an and numbers, which means that only these two features are the genre predictor for filter.

Meanwhile, based on Pennebaker (2011) female-preferential features, it is found that diary genre uses personal pronouns, verbs, negation, certainty words, and hedge phrases more than filter genre does. However, only two features shows significant p-values; personal pronoun and hedge phrases. In other words, only these two features are the genre predictor for diary. Furthermore, based on Pennebaker's male-preferential features, it isfound that "filter" uses number and preposition more than "diary", but

only number has significant p-value. It means that number is the predictor for genre filter.

For weblog author gender, based on Argamon et al's female-preferential features, it is found that female uses first-person singular, first person plural, third-person singular, and third-person plural more than male does. Furthermore, all features are significant but third-person Male. Therefore, we can conclude that personal pronoun is female predictor, except for third-person Male. Male authors use a/an and number more than female authors do. Differently, in this study, other male predictors are used by female. Female authors use demonstrative and its more than male authors do. For the significant p-values, only a/an and number show significance, meaning that these two features are the gender predictor for male.

In addition, based on Pennebaker (2011) female-preferential features, it is found that female authors use personal pronouns, verbs, negation, certainty words, and hedge phrases more than male-authors do. Moreover, all p-values are significant which means that all these features are gender predictor for female. For (Pennebaker, 2011) male-preferential features, it is found that male uses number and preposition more than female, but only preposition has significant p-value. It means that preposition is the gender predictor for male.

This study employed the preferential features found by previous studies (Argomon et al and Pennebaker's) and found that not all those features were found significant for the context this study was taken. This is due to the different characteristic of the authors' background that were coming from different language environment, which is in English as Foreign Language (EFL) context. Therefore, it is suggested for the future research to find out the preferential features in this EFL context itself without referring to the previous preference or predictor. From this, the similar findings of preference can strengthen the previous findings in language use preference, and the different preference can enrich the variation of language use preference, especially in EFL context.

In addition, to obtain the more statistically convincing results, more data is preferably needed. Therefore, the future research can include more text entries to analyze. Also, the language use preference can be obtained not only from weblog, but also from other modes of CMC. By studying other modes of CMC, the richer findings can be found to draw the pattern of language preference better.

ACKNOWLEDGMENTS

REFERENCES

Androutsopoulos, J. (2006). *Introduction : Sociolinguistics and computer-mediated communication*. 419–438.

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text*, *23*(3), 321–346. https://doi.org/10.1515/text.2003.014

Bodomo, A. B. (2010). Computer-Mediated Communication for Linguistics and Literacy. In *Computer-Mediated Communication for Linguistics and Literacy*. https://doi.org/10.4018/978-1-60566-868-0

Colley, A., & Todd, Z. (2002). Gender-linked differences in the style and content of e-mails to friends. *Journal of Language and Social Psychology*, *21*(4), 380-392+454. https://doi.org/10.1177/026192702237955

Domínguez-Rué, E. (2012). In Their Blooming Sixties: Aging as Awakening in Amanda Cross' The Imperfect Spy and The Puzzled Heart. *European Journal of Life Writing*, *1*, 1–21. https://doi.org/10.5463/ejlw.1.23

Ess, C., & Sudweeks, F. (Eds.). (n.d.). *Culture, Technology, Communication: Towards an intercultural global village*.

Herring, S. C. (2002). Computer-mediated communication on the Internet. *Annual Review of Information Science and Technology*, *36*, 109–168. https://doi.org/10.1002/aris.1440360104

Herring, S. C., & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics*, *10*(4), 439–459. https://doi.org/10.1111/j.1467-9841.2006.00287.x

Janssen, A., & Murachver, T. (2004). The role of gender in New Zealand literature: Comparisons across periods and styles of writing. *Journal of Language and Social Psychology*, *23*(2), 180–203. https://doi.org/10.1177/0261927X04263822

Koppel, M., Argamon, S., & Shimoni, A. R. (2004). Automatically Categorizing Written Texts by Author Gender. *Literary and Linguistic Computing*, *45*(1), 43. https://doi.org/10.1093/llc/17.4.401

Martin, J. R., & Rose, D. (2007). *Working with Discourse: Meaning Beyond the Clause*. Bloomsbury Publishing.

Mesthrie, R., Swann, J., Deumert, A., & Leap, W. L. (2009). *Introducing Sociolinguistics*. Edinburgh University Press.

Mooney, A., & Evans, B. (2015). Language, Society and Power. In *Language, Society and Power*. https://doi.org/10.4324/9781315733524

Pemberton, L., & Shurville, S. (2000). *Words on the web : computer mediated communication*.

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. Bloomsbury Press.

Puspita, D. (2019a). Error Analysis on Learners' Interlanguage and Intralanguage: A Case Study of Two Adolescent Students. *Teknosastik*, *17*(2), 12. https://doi.org/10.33365/ts.v17i2.321

Puspita, D. (2019b). Corpus Based Study: Students' Lexical Coverage Through Business Plan Report Writing. *2ND INTERNATIONAL CONFERENCE ON ENGLISH LANGUAGE TEACHING AND LEARNING*, *November*, 25–38.

Puspita, D. (2019c). *International Conference on English Language Teaching and Learning 2018. November*, 25–38.

Puspita, D., & Pranoto, B. E. (2021). The attitude of Japanese newspapers in narrating disaster events: Appraisal in critical discourse study. *Studies in English Language and Education*, *8*(2), 796–817. https://doi.org/10.24815/siele.v8i2.18368

Samar, R. G., & Shiazizadeh, M. (2010). *Gender-preferential Linguistic Elements in Applied Linguistics Research Papers: Partial Evaluation of a Model of Gendered Language*. *3*(1), 70–84.

Sari, B. N., & Gulö, I. (2019). Observing Grammatical Collocation in Students' Writings. *Teknosastik*, *17*(2), 25. https://doi.org/10.33365/ts.v17i2.325

Suprayogi, & Pranoto, B. E. (2019). Konstruksi Wacana Kasus Pembegalan di Portal Berita Tribun Lampung. In E. Citraresmana & T. Ampera (Eds.), *MANUSIA, BAHASA, DAN TEKNOLOGI* (pp. 12–32). Unpad Press.

Suprayogi, S. (2019). Javanese Varieties in Pringsewu Regency and Their Origins. *Teknosastik*, *17*(1), 7. https://doi.org/10.33365/ts.v17i1.243

Thomson, R., & Murachver, T. (2001). Predicting gender from electronic discourse. *British Journal of Social Psychology*, *40*(2), 193–208. https://doi.org/10.1348/014466601164812

Tillery, D. (2005). The plain style in the seventeenth century: Gender and the history of scientific discourse. *Journal of Technical Writing and Communication*, *35*(3), 273–289. https://doi.org/10.2190/MRQQ-K2U6-LTQU-0X56

Triyono, S. (2003). Satuan Lingual Penanda Gender. *Humaniora Volume XV*, *15*(3), 317–326. https://doi.org/10.22146/jh.v15i3.798

.