

Gender-preferential language use in L1 and L2 argumentative essays? Evidence against lists of ‘gendered’ language features

Sulistya Ningrum¹ and Peter Crosthwaite^{2*}

¹*PJ Zoetmulder Building, G102 Level 1, Faculty of Cultural Sciences, Universitas Gadjah Mada, Jl. Nusantara 1, Bulaksumur, Yogyakarta 55281, Indonesia*

²*Applied Linguistics, School of Languages and Cultures, University of Queensland, 510 Gordon Greenwood Building, St Lucia, Queensland 4072, Australia*

ABSTRACT

This study identifies and compares the gender-preferential language features present in the argumentative writing of L1 Indonesian and Indonesian L2 English learners. The data is comprised of 80 English argumentative essays sampled from the International Corpus Network of Asian Learners of English (ICNALE, Ishikawa, 2011) and a comparative corpus of 80 L1 Indonesian argumentative essays collected online from Indonesian university students, both equally divided by gender. Comparison of the data was performed through quantitative analysis of three supposed ‘male-preferential’ features and seventeen ‘female-preferential’ features between the male- and female-produced corpora in L1 and L2 writing. This study investigated (1) the extent of variation in the use of ‘gendered language features’ between male and female-produced L1 and L2 texts; (2) whether the use of male/female ‘gendered-language features’ across male/female produced L1/L2 texts match their suggested gender preference, and (3) to what extent L1’s preference for ‘gender language features’ affects male and female learners’ use of such language in L2. The results suggest the majority of supposed gender-preferential features were not significantly different across male/female produced texts, indicating that argumentative essays may be gender-neutral to a certain extent. This study also revealed that L1 preference of gendered language forms does not determine their preferences in the L2. In conclusion, male and female students adopt similar linguistic features to express their arguments. We may claim that gender language forms are not fixed and absolute in academic discourse because instructive texts tend to have a set model to fulfil the pedagogical criteria.

Keywords: Academic writing; Contrastive Interlanguage Analysis (CIA); gender; Indonesian L2 English

First Received:

19 April 2019

Revised:

13 September 2019

Accepted:

12 November 2019

Final Proof Received:

30 May 2020

Published:

31 May 2020

How to cite (in APA style):

Ningrum, S., & Crosthwaite, P. (2020). Gender-preferential language use in L1 and L2 argumentative essays? Evidence against lists of ‘gendered’ language features. *Indonesian Journal of Applied Linguistics*, 10(1), 218-233. <https://doi.org/10.17509/ijal.v10i1.25100>

INTRODUCTION

We, as individuals, live in a society where our behaviour is often understood and interpreted based on gender. This interpretation is mediated through a combination of social, cultural, political, and economic influences (Talbot, 1998). Considering the strong and dynamic connection between gender and

human behaviour, gender is a crucial variable in the study of many disciplines including psychology, sociology, arts, anthropology, as well as studies of language (Krijnen & Van Bauwel, 2015). For example, sociolinguists have defined gender as the combination of socially constructed human attributes and identities, whereas sex refers to the biological

* Corresponding Author
Email: p.cros@uq.edu.au

and genetic differences between males and females (Eckert & McConnell-Ginet, 1995; Jule, 2017; Lakoff, 1972; Oakley, 1972; Sunderland, 2000; Talbot, 1998).

The study of language and gender was not prevalent until the breakthrough work of Lakoff (1972). Since then, several studies have emerged in this area, discussing gendered language in a variety of contexts (e.g., Gilligan, 1982; Maltz & Borker, 1982; Swacker, 1975; Tannen, 1990). The debate regarding gendered language use mainly centred around Lakoff's (1972) claims that (1) males and females use different forms of language and (2) the differences in this use are the result of male dominance. Approaches to the investigation of gender in language studies are therefore generally separated into the *dominance* and the *difference* approaches, where the former focuses on issues of equality and the latter explores the diversity of language use among men and women as well as the tolerances for such diversity (Eckert & McConnell-Ginet, 2013). Under the *difference* approach, previous research has sought to identify 'gendered' language features by analysing language variation in texts and speech produced across the genders. Gender-linked language differences are characterised through variation in the use of lexical, syntactic, structural, and content-specific features (Argamon et al., 2003; Koppel et al., 2002; Zheng et al., 2006). Researchers have used various terms in addressing this variation across genders including *gender-preferential features* (Samar & Shirazizadeh, 2010), *gender-based distinctions* (Baron, 2004), *gender-related language style* (Colley & Todd, 2002), *gender-typical style* (Rubin & Greene, 1992), *gender-based writing styles* (Argamon et al., 2003), and *gender-specific language characteristics* (Sarawgi et al., 2011). Despite variation in their titles, each refers to the notion that certain language features are used extensively and/or exclusively by either men or women when communicating with others. In this research, the term *gender-preferential features* (Samar & Shirazizadeh, 2010) will hereafter be used to refer to this distinction.

Linguists have categorised the central contrast of language divergence between males and females into two main classifications. Overall, female language is typically associated with an affiliative approach to language use, while male language is heavily characterised by an assertive approach. An affiliative approach suggests women tend to engage and interact positively with their audience (Flynn, 1988; Leaper & Ayres, 2007; Rubin & Green, 1992). Opposite to males, females also tend to focus on managing their relationship with their audience by positively acknowledging the position of others. For example, Leaper and Ayres (2007) listed a range of affiliative attributes, including offering support, a preference for agreement, and a willingness to

recognise others' contributions more than seen in male discourse.

On the other hand, the key functions of an assertive approach are giving directive statements, delivering information, as well as disapproving and criticising others' views (Leaper & Ayres, 2007). This approach is claimed to expand males' power to intervene directly and objectively. Furumo and Pearson (2007) also suggested that males tend to use more task-oriented and denotative commands about what others should do. Other studies (e.g. Farrell, 1979; Roen & Johnson, 1992; Taylor, 1978) supported this view, concluding that male language tends to be competitive, antagonistic, and aggressive, while female language is more cooperative and submissive.

In classifying these elements, Biber (1995) outlined an *involvement-informational dimension* that refers to the differences across male and female language use. Females are claimed to adopt the use of language features indicative of participatory involvement with their audience, including *egocentric sequences* (e.g., in my opinion, I believe) and *modal adjuncts* (e.g., maybe, hopefully). Female language has also been claimed to feature heavy use of *pronouns* (Argamon et al., 2003; Colley & Todd, 2002; Koppel et al., 2002) and *tag questions* (Baron, 2004; Sterkel, 1988), suggesting females tend to get personally involved in the situation they are discussing, or wish to directly interact with their audience. This is also claimed to be marked by the extensive use of *intensifiers*, e.g., *strongly*, *really*, and *very* (Mulac & Lundell, 1994; Rubin & Greene, 1992; Sterkel, 1988), *affective markers*, e.g., *excited* and *anxious* (Baron, 2004; Colley and Todd, 2002; Mulac & Lundell, 1994), and *diminutives*, e.g., *a kitty* for a cat and *veggie* for vegetables (Baron, 2004). Various studies (Baron, 2004; Koppel et al., 2002; Lakoff, 1973; Mulac & Lundell, 1994; Rubin & Greene, 1992) have also suggested that female language is strongly characterised by the extensive use of *hedges* (e.g., *somewhat*, *probably*), *perceptual verbs* (e.g., *seems*, *looks*), *adversative connectives* (e.g., *but*, *otherwise*), *auxiliaries of possibility* (e.g., *could*, *may*), *qualifiers* (e.g., *nearly*, *kind of*), and *conjunctions* (e.g., *and*, *but*). These are claimed to reveal an ambience of uncertainty and uneasiness in female language use as a result of perceived male-dominated academic fields. In a study, Rubin and Greene (1992) suggested a female gender-preferential language coding scheme based on previous studies (e.g., Flynn, 1988; Hiatt, 1977; Hunter et al., 1988; Rubin & Nelson, 1983; Scates, 1981). Table 1 shows the classification of female linguistic features as suggested by Rubin and Greene (1992), which are to be used in the investigation into gender-preferential features in this study.

On the other hand, males are claimed to exclude such features in their production in favour of 'informational' characteristics indicative of the

presentation of facts or information. These include *quantifiers*, e.g., *one, some, and more* (Koppel et al., 2002; Mulac & Lundell, 1994; Sterkel, 1988), *determiners*, e.g., *the, a, and an* (Argamon et al., 2003; Koppel et al., 2002) and *locatives*, e.g., *above, inside, and left* (Mulac et al., 1986; Mulac & Lundell, 1994) to directly present information or facts in their writing. Scates (1981) defined this objective approach as denotative, where the linguistic features used are intended to demonstrate explicit and precise

meanings. Although male writing is more likely to exclude expressive or emotional expression, *judgmental adjectives*, e.g., *distracting and bad-tempered* and *profanity*, e.g., *damn and hell* (Baron, 2004) are frequently used as a substitution for other, more female-oriented forms. Table 2 summarises the main distinguishing features of male language based on Koppel et al. (2002) and Mulac and Lundell (1994).

Table 1
Female-Preferential Linguistic Features

Categories	Examples
Egocentric sequences	In my opinion, I think, I believe
Refusals	I am not sure, I do not know, I disagree
Illative connectives	Therefore, so, consequently
Adversative connectives	However, but, yet
Causal connectives	Because, since, in order to
Illustrators	For example, for instance, as an illustration
Additive connectives	And, also, with, together with
Temporal connectives	Next, after, lastly,
Conditional connectives	If, as long as
First-Person pronouns	I, me, we
Second-person pronouns	You, your
Intensifiers	A lot, quite, really
De-intensifiers	Just, only, not really
Proximals	About, around, nearly
Modal adjuncts	Maybe, hopefully, probably
Auxiliaries of possibility	Can, could, may
Perceptual verbs	Looks, seems, feels

Table 2
Male-Preferential Linguistic Features

Features	Examples
Quantifiers	Some, many, plenty
Locatives	Above, inside, in,
Determiners	A, the, that, an, any, other, another,

However, it is possible that these differences among male/female language users may be genre/register specific. Genre is likely to influence the use of gendered language since the author must use the language features appropriate for and constitutive of the target genre, which may limit the range of ‘gendered’ forms they can use (Sterkel, 1988; Swales, 1990). Numerous studies investigating gender-preferential features have been conducted in the context of literary works (e.g., Fischer-Stracke, 2010; Holmes, 1998; Stubbs, 2005) and orally-produced discourse (e.g., Furumo & Pearson, 2007; Hyde & Linn, 1988; Leaper & Ayres, 2007; Schirmer et al., 2005). An example of a contrastive study across registers is that of Argamon et al. (2003) who explored the role of gender across non-fiction and fiction texts in the British National Corpus. The results showed that gender-preferential features in fiction documents were more prevalent compared to the non-fiction documents in the corpus. In a study exploring gender-related structural and rhetorical styles involving 100 fiction and non-fiction books, Hiatt (1976) also found that female fiction authors

used more “feminine verbs” associated with feelings, perceptions, and emotions than male fiction authors. In addition, gendered language forms were more likely to be observed in fiction than in non-fiction. Gendered language preferences are believed to more frequently be exhibited in genres such as literature, where the author is afforded the space to express themselves freely. This has been found in drama texts (Culpeper, 2009), novels (Fischer-Stracke, 2009, 2010), novellas (Stubbs, 2005), and poetry (Enkvist, 1964).

However, in academic argumentative writing, the use of gendered language forms appears to be reduced due to the formality and standardised structures of this register, where both males and females are restricted to the same standards (Mulac & Lundell, 1994). This is also seen in Smeltzer and Werbel’s (1986) study exploring samples of business texts. This study correlated differences in the language features used across male and female writers with that of writing quality, finding no significant differences among the genders. Likewise, Sterkel (1988) conducted a study investigating

twenty gender-linked text attributes, including qualifiers, superlatives, politeness words, and a coding scheme of direct/indirectness in business letters. Male and female authors did not differ significantly in their use of any of the twenty language features analysed.

Moreover, while there is a dearth of studies on gendered language features in academic writing, there are fewer still for L2 academic writing. Previous investigations of academic writing and gender features have been typically limited to monolingual sources, where researchers have focused on scientific articles (Argamon et al., 2003; Sarawgi et al., 2011; Koppel et al., 2002), essays (Engelhard et al., 1992; Jones & Myhill, 2007; Mulac & Lundell, 1994; Rubin and Greene, 1992), and web blogs (Sarawgi et al., 2011). Previous studies comparing the L1 and L2 use of gendered language features are as yet rare. Studies that have been done include Samar and Shirazizadeh (2010). They found that the role of gender was more evident in the authors' native language rather than in their second language, as L2 learners lacked the L2 vocabulary to express their stance in the manner predicted by the writers' gender. Argamon et al. (2003), exploring the British National Corpus, also found that gendered language is more noticeable in L1 than L2 production. However, there is still a great need for contrastive L1/L2 studies on gendered language use, particularly covering academic writing.

This proposed study explores the use of gender-preferential features in Indonesian L1 and L2 English argumentative academic writing under the Contrastive Interlanguage Analysis (CIA²) framework (Granger, 2015). Argumentative academic writing represents a dynamic use of language representing real-life experiences while at the same time offering assertions and evaluations of the presented evidence. This type of writing allows the writers to express their interpersonal voice in providing argumentative viewpoints and building mutual consensus with the readers (Hyland, 2005). The question remains as to whether this academic,

interpersonal voice stifles the use of gender-preferential language features by gender across L1 and L2 texts. Moreover, it remains to be seen whether (in fact) the notion of gender-preferential language features can be supported by corpus evidence. This study was, therefore, conducted to find out the extent of variation in the use of 'gendered language features' between male and female-produced L1 and L2 texts. Then, it also seeks whether the use of male/female 'gendered-language features' across male/female produced L1/L2 texts match their suggested gender preference. Finally, it determines to what extent L1's preference for 'gender language features' affects male and female learners' use of such language in L2.

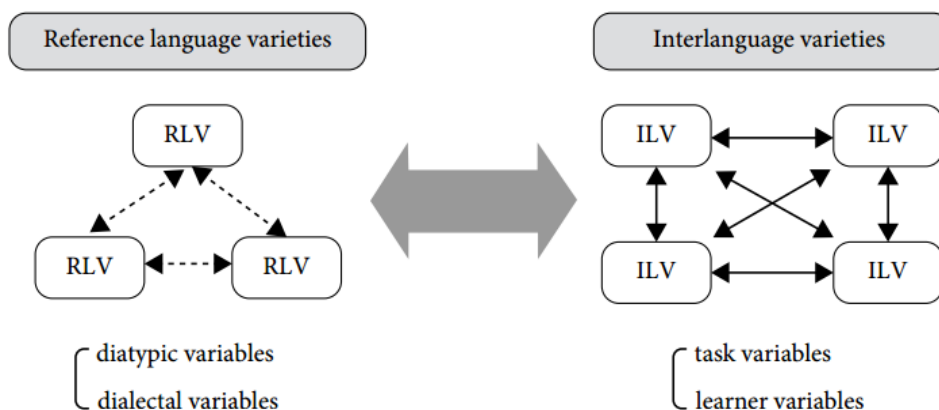
METHOD

In this study, the use of gendered language across the seventeen categories of female linguistic features by Rubin and Greene (1992) and three categories of male linguistic features by Koppel et al. (2002) and Mulac and Lundell (1994) are compared across two corpora of academic essays taken from L1 Indonesian and L1 Indonesian L2 English learners respectively. In particular, this research seeks to test the validity of these gender-preferential lists within academic writing and across gender and L1/L2 dimensions.

Contrastive Interlanguage Analysis

Contrastive Interlanguage Analysis (CIA) was proposed by Granger (1996) and refined in Granger (2015) as a methodology used to compare distinctive aspects of language use through language corpora. CIA studies have focused on native language vs. interlanguage varieties (e.g. Breckle & Zinsmeister, 2012; Chen, 2010; Hyland & Milton, 1997) as well as interlanguage vs. interlanguage varieties (e.g. Snape, 2008; Crosthwaite, 2016). The linguistic phenomena investigated under a CIA framework can cover grammatical and/or lexical analyses (Granger, 2015). The most recent version of the CIA framework is shown in Figure 1.

Figure 1
CIA² (Granger, 2015)



A previous version of CIA (Granger, 1996) sought only to compare L1/L2 and L2/L2 texts but was revised in Granger (2015) to focus on ‘varieties’, after claims that CIA resulted in the ‘comparative fallacy’ (Bley-Vroman, 1989). The latest iteration eschews an L1/L2 distinction, focusing instead on Reference Language Varieties (RLV) and Interlanguage Varieties (ILV). RLVs and ILVs can take many forms, e.g. novice L1 writers can be realised as an ILV, with professional L2 writers as an RLV. By carefully controlling for task and learner variables, a meaningful comparison of corpus data produced by different language groups can be conducted across dialectic and diatypic variables, in this case, argumentative academic essays. Our RLVs in this study include male- and female-produced essays from L1 Indonesian, while our ILVs in this study include male- and female-produced essays from L1 Indonesian L2 English learners. However, due to differences in the presence/frequency of particular grammatical categories between Indonesian/English (e.g. use of determiners is far more frequent in English), we do not directly compare ILV and RLV corpora quantitatively, instead of inferring ILV/RLV differences from our interpretation of the separate analyses.

Corpus data – ILV

The L1 Indonesian L2 English ILV data was sourced from the International Corpus Network of Asian Learners of English (ICNALE, Ishikawa, 2011), a freely available downloadable collection of L2 learners’ writing and speaking production compiled to facilitate CIA research (Ishikawa, 2011). The ICNALE written data is comprised of argumentative essays produced by both males and females across only two topics:

1. *It is important for college students to have a part-time job (henceforth PTJ).*
2. *Smoking should be completely banned at all the restaurants in the country (henceforth SMK).*

ICNALE texts are classified according to L2 learners’ English proficiency levels as determined by Nation and Beglar’s (2007) Vocabulary Size Test (VST) and the results of standardised tests including IELTS, TOEFL, or TOEIC. These data are combined to provide an approximation of L2 proficiency according to the Common European Framework of Reference (CEFR) bands (Council of Europe, 2001). The L1 Indonesian L2 English section of the ICNALE corpus is comprised of 200 EFL learners (107 males and 93 females) across four CEFR proficiencies.

For the purposes of the present study, our data sample is taken only from the ICNALE’s B1-2 (corresponding to a mid-point between CEFR B1-B2 levels). The total number of texts written under the PTJ prompt numbered 48 texts from male writers and 35 texts from female writers, with an equal number from both groups for texts produced under the SMK prompt. Texts from lower levels were not considered as the texts did not contain enough of the gendered language features for analysis, while there were only six texts at the highest ICNALE B2 level from writers from L2 Indonesian language backgrounds. Due to the poor quality of some of the ICNALE texts (e.g. some texts appeared to show the use of translation software, or were duplicates), the researchers hand-picked 80 essays using purposive sampling across both PTJ/SMK prompts, with 40 produced by male writers and 40 by female writers. In total, the observed data involved 10,666 and 11,144 tokens from the male and female groups, respectively.

Corpus data – RLV

To generate an equivalent L1 Indonesian data set as an RLV, 40 Indonesian EFL students (20 females and 20 males) were recruited using random sampling from an undergraduate academic discussion course at a university in Indonesia. All were native L1 Indonesian speakers. The L1 Indonesian texts were collected online under the same conditions as that of the ICNALE data to ensure comparability. The two ICNALE writing prompts (PTJ/SMK) were translated into Indonesian, and participants were informed that each essay should be comprised of between 200-300 words and meet the requirement of a good argumentative essay with a clear thesis and supporting statements. The sample of writing collected totalled 80 Indonesian argumentative essays with a 50/50 split across PTJ/SMK prompts. With approximately 250 words submitted for each essay, in total, the data totalled 9,605 and 9,268 tokens from the male and female groups, respectively.

The overall structure of corpora used in this study is summarised for the reader in Figure 2.

Analysis

The researchers generated a set of wordlists of gender-preferential language features by expanding the existing lists created by Koppel et al. (2002), Mulac and Lundell (1994) and Rubin and Greene (1992) in both English and Indonesian languages (Tables 3-6). The examples for each category were checked by two native speakers from each language background to ensure accuracy and completeness.

Figure 2
Research Data Mapping

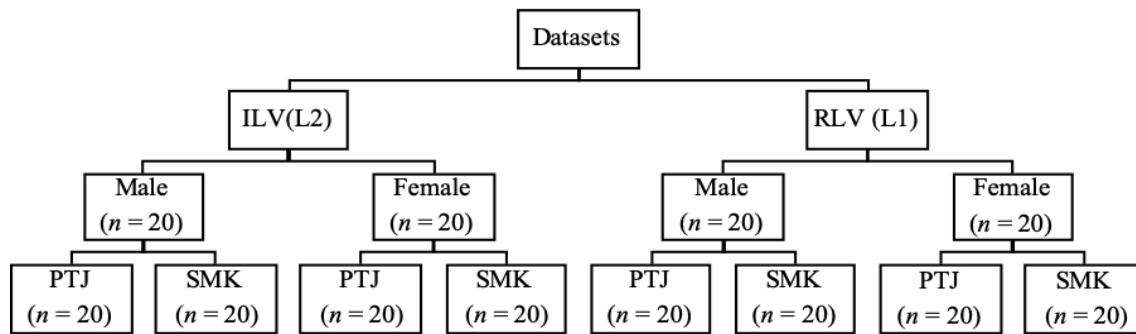


Table 3
Wordlist of Male-Preferential Features in English Language

Features	Examples
Quantifiers	Some, many, plenty, heap, load, ton, both, each, either, few, neither, several, couple, bit, all, a lot of, a number of, a plethora of, enough, sufficient, no lack of, lots of, adequate, as much as, ample, abundant
Locatives	Above, inside, in, at, on, near, there, here, below, within, centre, middle, corner, front, around
Determiners	A, the, that, an, any, other, another

Table 4
Wordlist of Female-Preferential Features in English Language

Features	Examples
Egocentric Sequences	In my view, in my standpoint, in my point of view, in my perspective, in my opinion, I would say, I would argue, I think, I suspect, I suppose, I reckon, I presume, I personally think, I guess, I expect, I contemplate, I conclude, I believe, I assume, I anticipate, from my standpoint, from my point of view, from my perspective, from my personal standpoint, from my observation, according to my perspective, according to my opinion, according to me
Refusals	I am not sure, I do not know, I cannot rightly say, I do not think, I do not agree, I disagree, I disbelief, I do not believe
Illative connectives	Therefore, so, consequently, as a result, as a consequence, hence, thus, accordingly
Adversative connectives	However, but, yet, otherwise, nevertheless, nonetheless, still, though, although, even so, despite that, in spite of that, anyway, anyhow, notwithstanding
Causal Connectives	Because, since, in order to
Illustrators	For example, for instance, as an illustration, such as, to illustrate, namely, like
Additive Connectives	And, also, with, together with, along with, as well as, in addition, including, too, besides, furthermore, moreover, plus
Temporal Connectives	Next, after, lastly, first, afterwards, subsequently, thereafter, thereupon, then
Conditional Connectives	If, as long as
First-Person Pronouns	I, me, my, we, us, our
Second-person Pronouns	You, your
Intensifiers	A lot, quite, really, very, extremely, at all, ever, too, so
De-Intensifiers	Just, only, not really, rather, approximately, roughly
Proximals	About, around, nearly, roundabout, thereabouts, more or less, close to, almost
Modal Adjuncts	Maybe, hopefully, probably, possibly, perhaps, conceivably, feasibly, likely
Auxiliaries of possibility	Can, could, may, might, will, would, shall, should, must
Perceptual verbs	Look, seem, sound, feel, taste, hear, listen, watch, smell, looks, seems, sound, feels, tastes, hears, listens, watches, smells, looked, seemed, sounded, tasted, heard, listened, watched, smelled

Table 5
Wordlist of Male-Preferential Features in Indonesian Language

Features	Examples
Quantifiers	<i>Beberapa, keduanya, masing-masing, sedikit, semua, sejumlah, kebanyakan, cukup, tidak ada, jumlah, banyak, sebanyak, berlimpah, kuantitas, angka, seluruh, tidak ada, tak ada, kekurangan, kurang, kurangnya, berlebih</i> [Some, both, each, a little, all, a number, most, enough, nothing, total, many, as much, abundant, quantity, number, whole, nothing, none, deficiency, less, lack of, excessive]
Locatives	<i>Di, ke, atas, dalam, dekat, sana, sini, situ, bawah, tengah, sudut, depan, belakang, sekitar</i> [In, to, above, in, near, there, here, there, under, middle, corner, front, back, around]
Determiners	<i>Para, sebuah, suatu</i> [the, a, a]

Table 6
Wordlist of Female-Preferential Features in Indonesian Language

Features	Key Examples
Egocentric Sequences	<i>Saya berasumsi, saya simpulkan, saya pikir, saya percaya, saya menyimpulkan, saya mengira, saya menduga, saya kira, saya duga, saya beropini, saya berargumen, menurut saya, menurut sudut pandang saya, menurut pendapat saya, dalam pandangan saya, dalam sudut pandang saya</i> [I assume, I conclude, I think, I believe, I conclude, I suppose, I guess, I think, I argue, In my opinion, I argue, in my opinion, In my standpoint, based on my opinion, in my view, in my point of view]
Refusals	<i>Saya tidak yakin, saya tidak tahu, saya tidak bisa memastikan, saya tidak setuju, saya tidak percaya</i> [I am not sure, I do not know, I cannot be sure, I disagree, I do not believe]
Illative connectives	<i>Oleh karena itu, demikian, akibatnya, konsekuensinya, karenanya, dengan demikian, maka, oleh sebab itu</i> [Therefore, thus, as a result, as a consequence, so, hence, then, for that reason]
Adversative connectives	<i>Namun, tetapi, sebaliknya, meskipun demikian, meskipun begitu, walaupun begitu, walaupun demikian, terlepas dari itu, akan tetapi, bagaimanapun</i> [But, yet, on the contrary, in spite of this, however, despite all that, despite that, regardless, nevertheless, nonetheless]
Causal Connectives	<i>Karena, sebab, dikarenakan, disebabkan oleh, agar, supaya</i> [because, since, as, cause, in order to, to]
Illustrators	<i>Misalnya, seperti, yaitu, contohnya, contoh, misal</i> [For example, such as, that is, for instance, for example, like]
Additive Connectives	<i>Dan, juga, dengan, bersama, serta, di samping itu, termasuk, selain itu, ditambah, dan sebagainya, dan lain-lain</i> [And, also, with, together with, as well, besides, including, in addition to, as well as, and so on, and so forth]
Temporal Connectives	<i>Berikutnya, setelah, akhirnya, sesudah, kemudian, berikut, setelahnya</i> [Next, then, finally, after that, afterward, subsequently, later]
Conditional Connectives	<i>Jika, selama, kalau, apabila, jikalau, semisal</i> [If, as long as, in the event that, given that, on the condition that, assuming that]
First-Person Pronouns	<i>Saya, aku, kami, kita</i> [I, me, we, us]
Second-Person Pronouns	<i>Anda, kamu</i> [You, you]
Intensifiers	<i>Sangat, terlalu, begitu</i> [Very, too, extremely]
De-Intensifiers	<i>Hanya, tidak benar-benar, agak, kira-kira</i> [Only, not really, slightly, approximately]
Proximals	<i>Sekitar, hampir, kurang lebih, paling tidak, mencapai</i> [Around, almost, more or less, at least, nearly]
Modal Adjuncts	<i>Mungkin, semoga</i> [Maybe, hopefully]
Auxiliaries of possibility	<i>Akan, seharusnya, dapat, bisa</i> [Will, should, can, may]
Perceptual verbs	<i>terlihat, terdengar, terasa, dirasa, dilihat, didengar, rasanya, merasa, mendengar, melihat, dihirup, terhirup, menghirup, dirasakan, didengarkan, diperlihatkan</i> [seen, heard, felt, seem, looked, heard, taste, feel, hear, watch, inhaled, smelled, smell, felt, heard, shown]

To determine the frequencies of the features contained in Tables 3-6, the corpus query tool *CoQuery* (Version 0.10.0, Kunter, 2017) was used. Rather than searching for each word individually, *CoQuery* allows the user to derive the frequencies of all query terms simultaneously from a .csv file. For certain language features that may fulfil multiple functions (e.g. modals), the researchers double-checked the results for these by consulting concordance lines for each word/phrase and reducing the total frequency for each term identified as serving an alternative function, or where the function was not clear from the concordance output. The output frequency lists were exported to Excel sheets to facilitate further statistical analysis using the *Log-Likelihood Calculator* (Rayson & Garside, 2000) to determine whether any gender-preferential features or categories were significantly over- or under-used across ILVs and RLVs using the Log-likelihood Test

(G^2), with effect-size values calculated under the Effect Size for Log-Likelihood (ELL) criterion (Johnston et al., 2006).

FINDINGS

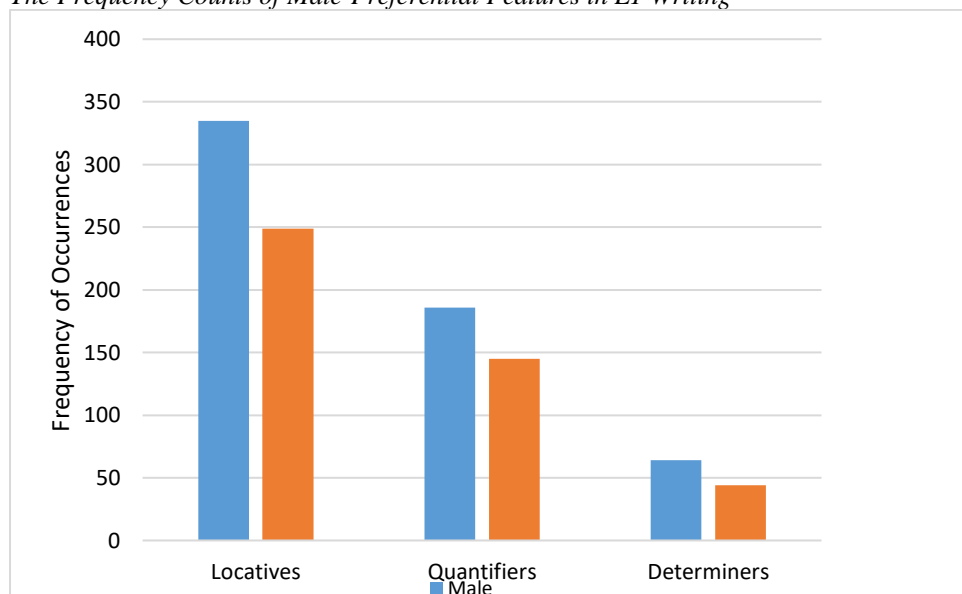
Gender-preferential features in Indonesian language (L1) writing

Male-preferential features

Referring to the works by Koppel et al. (2002) and Mulac and Lundell (1994), the male writing style is characterised by frequent use of ‘informational’ linguistic features including *determiners*, *locatives*, and *quantifiers*. This section presents the male-preferential features used by male and female authors in the L1 Indonesian argumentative essays (RLV). Male authors tended to use *locatives*, *quantifiers*, and *determiners* more frequently than female authors (Figure 3).

Figure 3

The Frequency Counts of Male-Preferential Features in L1 Writing



The data in Table 7 suggests that all three male-preferential feature categories were used more frequently by male writers, but only significantly so for *locatives* ($G^2 = 9.58$; $p < 0.01$). The differences found in other features were not significant as their

Log-likelihood values were lower than the significance threshold of 3.84. The effect size (ELL) shown for *locatives*, however, was not large (ELL = 0.00009).

Table 7

Male-Preferential Features in Indonesian Language (L1) Writing

Features	Male	Female	Over/under-use	Log-likelihood	ELL
Locatives	335	249	+	9.58**	0.0000900000
Quantifiers	186	145	+	3.62	0.0000400000
Determiners	64	44	+	2.99	0.0000400000

Note. *** $p < 0.0001$; critical value = 15.13
 ** $p < 0.01$; critical value = 6.63

*** $p < 0.001$; critical value = 10.83
 * $p < 0.05$; critical value = 3.84

As shown in Table 8, when looking at the use of male-preferential features at the word/phrase level, there were significant differences in the use of the

Indonesian quantifiers *angka* [number], *sejumlah* [a number of] and *jumlah* [total], as well as the determiners *sebuah* [a] and *suatu* [a], and the

locatives *di* [in] and *sekitar* [around]. Unpredictably, the locative *ke* [to] that was hypothesised to be more likely used by males was significantly more likely to

be used by females ($G^2 = 3.97$; $p < 0.05$; ELL = 0.0001).

Table 8

Selected Results of Words/Phrases Categorised as Male-Preferential Features in Indonesian Language (L1) Writing

Words	Male	Female	Over-/underused	Log-likelihood	ELL
<i>Angka</i> [Number]	17	0	+	22.97****	0.0005700000
<i>Sejumlah</i> [all]	12	1	+	11.65***	0.0004700000
<i>Jumlah</i> [total]	13	2	+	8.63**	0.0002300000
<i>Sebuah</i> [a]	9	1	+	7.08**	0.0002400000
<i>Di</i> [in]	228	169	+	6.82**	0.0000700000
<i>Sekitar</i> [around]	18	6	+	5.86*	0.0001300000
<i>Suatu</i> [a]	16	5	+	5.67*	0.0001300000
<i>Ke</i> [to]	5	13	-	3.97*	0.0001000000

Note. **** $p < 0.0001$; critical value = 15.13
 ** $p < 0.01$; critical value = 6.63

*** $p < 0.001$; critical value = 10.83
 * $p < 0.05$; critical value = 3.84

Female-preferential features

The ‘involved’ female-produced data has seventeen distinct categories as classified by Rubin and Greene (1992). Figure 4, which compares the frequency counts for female-preferential features in the L1 essays, indicates that most of the female-preferential features were used more frequently by females than male writers, including *additive connectives*, *second-person pronouns*, *modal adjuncts*, *egocentric*

sequences, *refusals*, *illustrators*, *auxiliaries of possibility*, *intensifiers*, *perceptual verbs*, *conditional connectives*, *illative connectives*, and *de-intensifiers*. The other five features that were hypothesised to be more frequently used by the female authors were relatively underused, including *first-person pronouns*, *proximals*, *causal connectives*, *temporal connectives*, and *adversative connectives*.

Figure 4

The Frequency Counts of Female-Preferential Features in L1 Writing

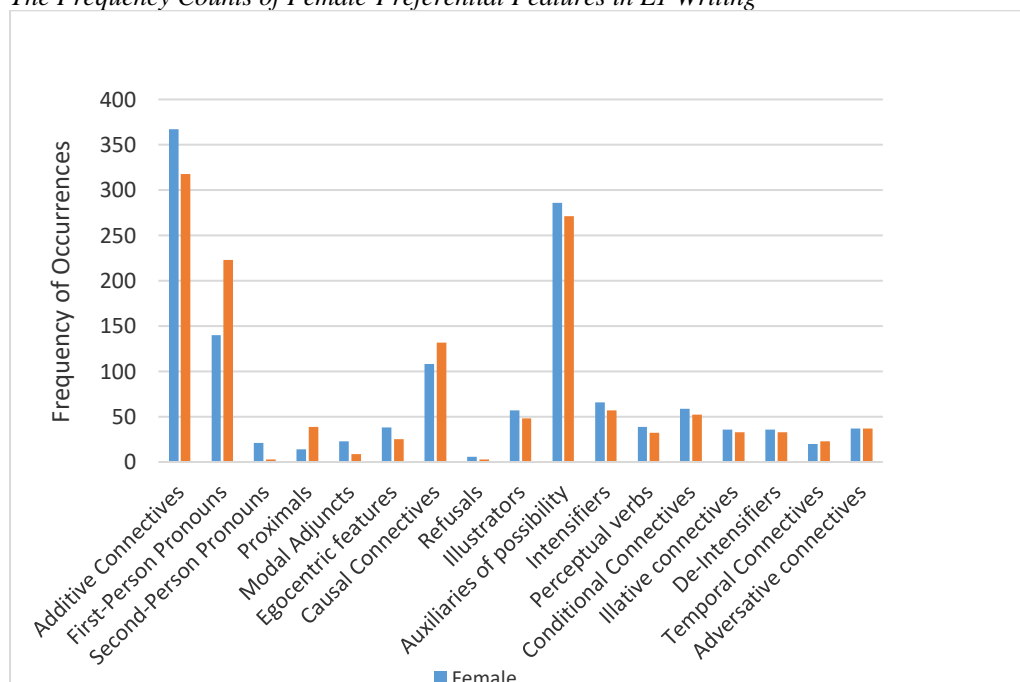


Table 9 shows that the use of *additive connectives*, *second-person pronouns*, and *modal adjuncts* are significantly linked to female writing. However, *first-person pronouns* and *proximals*—supposedly female-preferential features—were significantly more likely to be used by male authors.

Furthermore, no significant differences were observed across the genders for the other twelve categories, showing that there appears to be little impact of gender on the use of female-preferential features in argumentative writing.

Table 9
Female-Preferential Features in Indonesian Language (L1) Writing

Features	Female	Male	Over-/Underuse	Log-likelihood	ELL
Additive Connectives	367	318	+	34.56****	0.0003200000
First-person pronouns	140	223	-	16.30****	0.0001700000
Second-person pronouns	21	3	+	15.84****	0.0003400000
Proximals	14	39	-	11.40***	0.0001900000
Modal adjuncts	23	9	+	6.85**	0.0001300000
Egocentric features	38	25	+	3.19	0.0000500000
Causal connectives	108	132	-	1.62	0.0000200000
Refusals	6	3	+	1.13	0.0000400000
Illustrators	57	48	+	1.13	0.0000200000
Auxiliaries of possibility	286	271	+	1.12	0.0000100000
Intensifiers	66	57	+	1.02	0.0000100000
Perceptual verbs	39	32	+	0.96	0.0000100000
Conditional connectives	59	52	+	0.73	0.0000100000
Illative connectives	36	33	+	0.26	0.0000000000
De-intensifiers	36	33	+	0.26	0.0000000000
Temporal connectives	20	23	-	0.12	0.0000000000
Adversative connectives	37	37	-	0.02	0.0000000000

Note. **** $p < 0.0001$; critical value = 15.13
** $p < 0.01$; critical value = 6.63

*** $p < 0.001$; critical value = 10.83
* $p < 0.05$; critical value = 3.84

Table 10 summarises the results for female-preferential features at the word/phrase level. Six words/phrases were significantly more likely to be used by females (*Anda* [second-person pronouns], *mungkin* [modal adjuncts], *menurut saya* [egocentric sequences], *namun* [adversative connectives], *terdengar* [perceptual verbs], and *serta* [additive connectives]), while three words were significantly

more likely to be used by males (*kita* [first-person pronouns], *sekitar* [proximals], and *agar* [causal connectives]). There were no significant differences observed in the use of other seventy female-preferential words/phrases found in the corpus across the genders, while thirty-six words/phrases were absent from either sub-corpus.

Table 10
Selected Results of Words/Phrases Categorised as Female-Preferential Features in Indonesian Language (L1) Writing

Words	Female	Male	Over-/underuse	Log-likelihood	ELL
<i>Kita</i> [we]	15	82	-	48.56****	0.0006700000
<i>Anda</i> [you]	21	3	+	15.84****	0.0003400000
<i>Sekitar</i> [around]	12	36	-	11.72***	0.0002000000
<i>Agar</i> [in order to]	4	16	-	7.29**	0.0001700000
<i>Mungkin</i> [maybe]	23	9	+	6.85**	0.0001300000
<i>Menurut saya</i> [in my opinion]	33	18	+	5.03*	0.0000800000
<i>Namun</i> [but]	21	10	+	4.39*	0.0000900000
<i>Terdengar</i> [heard]	3	0	+	4.27*	0.0005800000
<i>Serta</i> [as well]	13	5	+	3.97*	0.0001000000

Note. **** $p < 0.0001$; critical value = 15.13
** $p < 0.01$; critical value = 6.63

*** $p < 0.001$; critical value = 10.83
* $p < 0.05$; critical value = 3.84

Gender-preferential features in English (L2) writing

Male-preferential features

Figure 5 and Table 11 compare the results of the occurrences of the three categories across male and female writing in the L2 English dataset.

As shown in Table 11, the Log-likelihood value for *determiners* was 4.84, showing a statistically significant difference at the level of $p < 0.05$. On another note, the use of *locatives* and *quantifiers* across both genders were not seen as significantly

different despite a higher raw frequency in the male data than the female data.

Table 12 describes which particular expressions (across *determiners*, *locatives* and *quantifiers*) were more likely to be used by male writers as compared with females. Due to limited space, only the words and phrases which exhibited significant differences in use across male/female subcorpora are provided. The data shows that out of the forty-eight expressions listed as male-preferential, only three were significantly overused by males, namely the locative *at*, the quantifier *many*, and the determiner *the*.

Figure 5
The Frequency Counts of Male-Preferential Features in L2 Writing

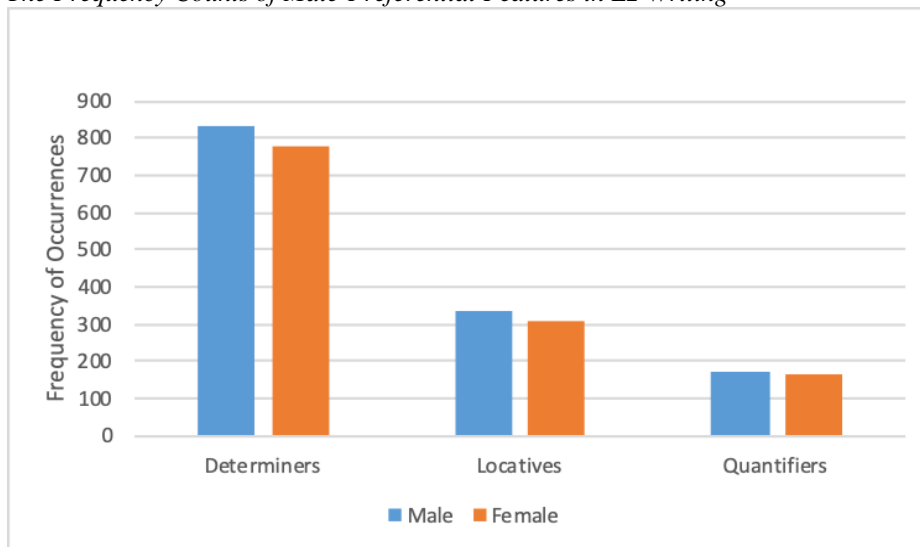


Table 11
Male-Preferential Features in English (L2) Writing

Features	Male	Female	Over-/underused	Log-likelihood	ELL
Determiners	831	778	+	4.84*	0.0000300000
Locatives	334	306	+	2.76	0.0000200000
Quantifiers	171	163	+	0.70	0.0000100000

Note. **** $p < 0.0001$; critical value = 15.13 *** $p < 0.001$; critical value = 10.83
 ** $p < 0.01$; critical value = 6.63 * $p < 0.05$; critical value = 3.84

Table 12
Selected Results of the Words Categorised as Male-Preferential Features in English (L2) Writing

Words	Male	Female	Over-/underused	Log-likelihood	ELL
At	63	33	+	10.9***	0.0001300000
Many	68	42	+	7.40**	0.0000900000
The	470	427	+	4.38*	0.0000300000

Note. **** $p < 0.0001$; critical value = 15.13 *** $p < 0.001$; critical value = 10.83
 ** $p < 0.01$; critical value = 6.63 * $p < 0.05$; critical value = 3.84

Female-preferential features

Figure 6 compares the frequency of occurrence of the female-preferential features across the male- and female-produced L2 essays. Ten features were seen to be more prevalent in the female-produced subcorpus, including *second-person pronouns*, *de-intensifiers*, *additive connectives*, *adversative connectives*, *temporal connectives*, *proximals*, *auxiliaries of possibility*, *conditional connectives*, *first-person pronouns*, and *egocentric sequences*. However, against predictions, the remaining seven features (*causal connectives*, *modal adjuncts*, *refusals*, *perceptual verbs*, *illustrators*, *intensifiers*, and *illative connectives*) occurred more often in the male-produced essays.

Table 13 depicts the Log-likelihood and ELL results across the male- and female-produced data for the use of female-preferential features. Only four features seem to be significantly more likely to be used by females; namely *second-person pronouns*, *de-intensifiers*, *additive connectives*, and *adversative*

connectives. There were no significant differences observed in the use of the other listed female-preferential features across the male- and female-produced data. Table 14 describes the over/underuse of individual female-preferential features across all categories in the male- and female-produced data. The data shows the words *you*, *just*, *your*, *still*, *therefore*, *also*, *I believe*, *however*, *might*, *first*, and *will* were significantly more likely to be used by female L2 writers, in descending order of significance. However, against predictions, the words auxiliary of possibility *must* and causal connective *because* were significantly more likely to be used by male L2 writers. There were no other significant differences in male/female use across other seventy-six female-preferential expressions that were present in the L2 corpus. Seventy-nine additional female-preferential expressions were not found in the L2 corpus and were therefore excluded from the statistical analysis.

Figure 6
The Frequency Counts of Female-Preferential Features in L2 Writing

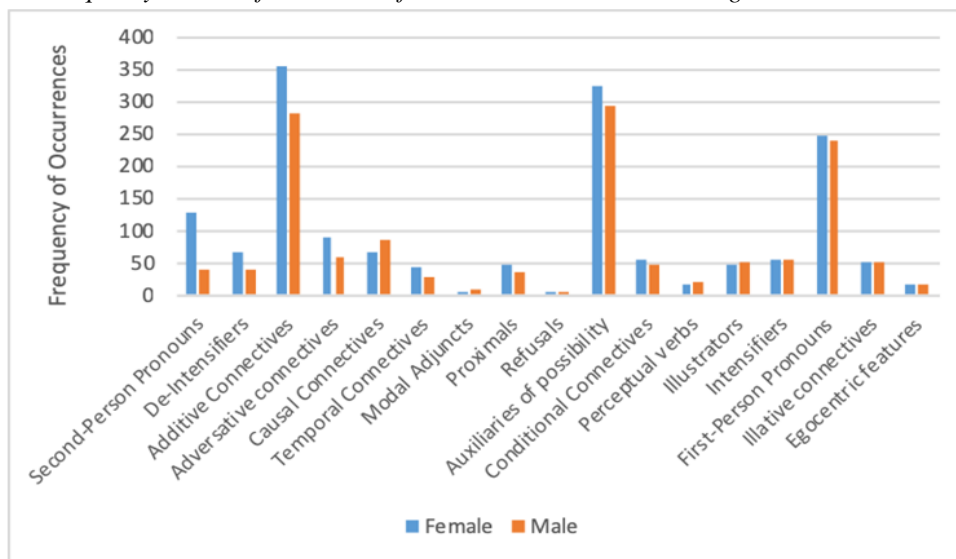


Table 13
Female-Preferential Features in English (L2) Writing

Features	Female	Male	Over-/underused	Log-likelihood	ELL
Second-person pronouns	127	38	+	46.83****	0.0004900000
De-intensifiers	68	38	+	7.34**	0.0000900000
Additive connectives	353	283	+	4.96*	0.0000400000
Adversative connectives	90	60	+	4.80*	0.0000500000
Causal connectives	67	86	-	3.27	0.0000300000
Temporal connectives	45	29	+	2.82	0.0000400000
Modal adjuncts	5	10	-	1.93	0.0000400000
Proximals	49	35	+	1.77	0.0000200000
Refusals	2	4	-	0.77	0.0000300000
Auxiliaries of possibility	323	293	+	0.44	0.0000000000
Conditional connectives	56	48	+	0.32	0.0000000000
Perceptual verbs	18	20	-	0.21	0.0000000000
Illustrators	49	51	-	0.18	0.0000000000
Intensifiers	54	55	-	0.11	0.0000000000
First-person pronouns	245	240	+	0.07	0.0000000000
Illative connectives	53	53	-	0.05	0.0000000000
Egocentric sequences	17	16	+	0.00	0.0000000000

Note. **** $p < 0.0001$; critical value = 15.13
** $p < 0.01$; critical value = 6.63

*** $p < 0.001$; critical value = 10.83
* $p < 0.05$; critical value = 3.84

Table 14
Selected Results of the Words/Phrases Categorised as Female-Preferential Features in English (L2) Writing

Words	Female	Male	Over-/underused	Log-likelihood	ELL
You	89	21	+	42.31****	0.0004900000
Just	20	3	+	13.34***	0.0002500000
Must	13	34	-	10.67**	0.0001600000
Your	38	17	+	7.33**	0.0001000000
Still	18	9	+	5.72*	0.0001100000
Therefore	4	0	+	5.37*	0.0003700000
Also	63	39	+	4.70*	0.0000600000
Because	60	81	-	4.13*	0.0000400000
I believe	3	0	+	4.03*	0.0004800000
However	3	0	+	4.03*	0.0004800000
Might	3	0	+	4.03*	0.0004800000
First	17	7	+	3.87*	0.0000700000
Will	87	60	+	3.87*	0.0000400000

Note. **** $p < 0.0001$; critical value = 15.13
** $p < 0.01$; critical value = 6.63

*** $p < 0.001$; critical value = 10.83
* $p < 0.05$; critical value = 3.84

DISCUSSION

What is the extent of variation in the use of 'gendered language features' between male and female-produced L1 and L2 texts?

The results of our analyses suggest that there are instances of significant variation in the use of 'gender-preferential' language features across male- and female-produced L1/L2 essays, although these are far less prevalent than was predicted by the literature. Of the so-called 'male-preferential' features, one out of three (*locatives*) exhibited significant differences across male- and female-produced writing in L1 Indonesian, with only determiners in the L2 writings. While this partly supports the findings of Koppel et al., (2002), Mulac and Lundell (1994), and Sterkel (1998) about the use of *determiners* as an indicator of male writing and Argamon et al. (2003) about *locatives*, overall, our data, in fact, indicates that male and female authors of argumentative essays tend to use 'informational' language features equally. This trend is also found at the word/phrase level, with only eight out of thirty-two word/phrases showing significant male/female differences in the L1 data, and only three out of thirty-three word/phrases in L2 data, indicating that male and female authors used the majority of the 'male-preferential features' at similar frequencies. However, the absence of any significant differences for *quantifiers* is in opposition to the outcomes of previous studies (Biber et al., 1998; Koppel et al., 2002; Mulac & Lundell, 1994; Scates, 1981; Sterkel, 1998; Swacker, 1975) that claim such forms are more likely to be found in male-produced language.

Regarding female-preferential features, out of seventeen female-preferential categories listed for both L1/L2, only three out of seventeen features showed significant male/female differences in use (*additive connectives*, *second-person pronouns*, and *modal adjuncts*), with only four in the L2 data (*second-person pronouns*, *de-intensifiers*, *additive connectives*, and *adversative connectives*). This trend is also seen at the word and phrase-level, where only 9 out of 79 features present in the corpus showed differences in male/female use in the L1 data, with 13 out of 79 in the L2 data. To some degree, the increased presence of *second-person pronouns*, *de-intensifiers*, *additive connectives*, *modal adjuncts*, and *adversative connectives* in a female language is in line with the studies conducted by Mulac and Lundell (1994) and Sterkel (1988). However, most of the features exhibited similar frequencies of use between men and women, both in L1 and L2, which is more in line with the findings of Sarawgi et al. (2011).

As the results for L1/L2 groups appear to be similar, we mainly attribute our findings to the genre in which the texts were produced. Our results indicate that argumentative essays tend to be gender-neutral to a certain extent, confirming the studies of Samar and Shirazizadeh (2010) and Mulac and Lundell

(1994). Of course, we would require a follow-up study comparing male/female-gendered language use across multiple genres to confirm this hypothesis.

Does the use of male/female 'gendered-language features' across male/female produced L1/L2 texts match their suggested gender preference?

As mentioned in the previous section, there was some positive evidence that certain gender-preferential features were in fact more likely to be used by their hypothesised gender group, as seen in the use of *locatives* (L1) and *determiners* (L2) by men, as well as forms including *second person pronouns*, *de-intensifiers*, *additive connectives*, etc. by women. There were no significant differences between male/female produced texts in L1 or L2 for the majority of 'gendered' linguistic categories and individual word/phrase features in the data.

In fact, certain 'female' preferential features were shown to be more frequently used by men in L1 Indonesian, and vice-versa. *ke*, a male-preferential *locative* (which translates to *to* in English) was found to be more frequently used by women at the $p < .05$ level. *First-person pronouns* and *proximals*—considered 'female-preferential'—were more likely to be used by men, while the expressions *kita* [we], *sekitar* [around], and *agar* [so that] were also more likely to be used by male writers. In L2 English, *must* and *because*, considered as female-preferential forms, were more likely to be used by male writers. These findings seriously call into question the reliability of any list of so-called 'gender-preferential' features if corpus data suggests that certain linguistic categories or forms are in fact more likely to be used by the *opposite* gender for which they have been categorised.

To what extent do L1 preferences for 'gendered language features' influence male and female learners' use of such language in L2?

This research has explored the use of gender-preferential features in Indonesian L1 and Indonesian L2 English varieties at the upper-intermediate level of proficiency. While we cannot directly statistically compare L1/L2 data due to differences in the presence of certain gendered language features in the grammars of the L1/L2, we can infer from the individual results that the hypothesised preferences for gendered language use were not more likely to be observed in the L1 Indonesian texts than they were in the L2. While male authors used *locatives* more frequently in L1, this was not seen in the L2 data, where *determiners* occurred more frequently in the male-produced texts. For the female-preferential features, *additive connectives*, *second-person pronouns*, and *modal adjuncts* were indeed more likely to be used by women than men in L1, but only *additive connectives* and *second-person pronouns* were used more by women in L2. *De-intensifiers* and *adversative connectives*, which were not overused by

female writers in L1, were more likely to be used by women in L2. Considering these findings, we conclude that the preference for gendered language features is generally not more prominent in L1 than in L2, unlike the findings of Argamon et al. (2003) and Samar and Shirazizadeh (2010).

CONCLUSION

The present study has investigated the presence of so-called 'gender-preferential language' in L1 Indonesian and L2 English argumentative essays. Motivated by the lack of investigation of gendered language features in academic texts, we determined that lists of such features need to be treated with caution, as their predicted use is not always realised in either L1 or L2 data, and there is little evidence that any L1 preferential use carries over to L2. We also (partially) conclude that argumentative essays appear to be 'gender-neutral' to some degree, in that this genre does not appear to provide writers with enough opportunity to demonstrate their gender-preferential identity through language.

Two main limitations observed in this study dealing with the sizes of the corpora and the wordlist used for comparison. Firstly, the sizes of the corpora from which to generalise the results of the statistical analysis were relatively small. As a result, a larger corpus would allow for more reliable statistical comparison, although large-scale Indonesian L1/L2 corpora are still rare. Secondly, both Indonesian and English wordlists were compiled by the 1st author. It is possible that certain other words related to specific categories have not been included in the wordlists, although two native speakers of both languages were used to confirm the coverage of the wordlists used for the present study. We invite other researchers working on L1/L2 texts produced by speakers of Indonesian, or other languages, to address the limitations of this study in future research.

REFERENCES

- Argamon, S., Koppel, M., Fine, J., & Shimsoni, R. (2003). Gender, genre, and writing style in formal written texts. *Interdisciplinary Journal for the Study of Discourse*, 23(3), 321-346. <https://doi.org/10.1515/text.2003.014>
- Baron, S. (2004). See you online: Gender issues in college student use of instant messaging. *Journal of Language and Social Psychology*, 23(4), 397-423. <https://doi.org/10.1177/0261927X04269585>
- Biber, D. (1995). *Dimensions of register variation: A cross-linguistic perspective*. Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, D. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In G. Susan & S. Jacquelyn (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-68). Cambridge University Press
- Breckle, M., & Zinsmeister, H. (2012). A corpus-based contrastive analysis of local coherence in L1 and L2 German. *Discourse and Dialogue*, 7(2), 235-250. <https://pdfs.semanticscholar.org/8dd8/ad2cfc18d8e8781165983c0a9ae791a125b2.pdf>
- Chen, I. (2010). Contrastive learner corpus analysis of epistemic modality and interlanguage pragmatic competence in L2 writing. *Arizona working papers in SLA & teaching*, 17(2), 27-51. <https://journals.uair.arizona.edu/index.php/AZSLAT/article/view/21239>
- 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. <https://doi.org/10.1177/026192702237955>
- Council of Europe. (2001). *Common European framework of reference for languages*. University Cambridge Press.
- Crosthwaite, P. (2016). L2 English article use by L1 speakers of article-less languages. *International Journal of Learner Corpus Research*, 2(1), 68-100. <https://doi.org/10.1075/ijlcr.2.1.03cro>
- Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, 14(1), 29-59. <https://doi.org/10.1075/ijcl.14.1.03cul>
- Eckert, P., & McConnell-Ginet, S. (1995). Constructing meaning, constructing selves: Snapshots of language, gender, and class from Belten High. In S. Mills (Ed.) *Language and gender: Interdisciplinary perspectives* (pp. 469-507). Routledge.
- Eckert, P., & McConnell-Ginet, S. (2013). *Language and gender*. Cambridge University Press.
- Engelhard, G., Gordon, B., & Gabrielson, S. (1992). The influences of mode of discourse, experiential demand, and gender on the quality of student writing. *Research in the Teaching of English*, 26(3), 15-336.
- Enkvist, E. (1964). On defining style. In E. Enkvist, J. Spencer, & M. Gregory (Eds.), *Linguistics and style* (pp. 1-56). Oxford University Press.
- Farrell, T. J. (1979). The female and male modes of rhetoric. *College English*, 40(8), 909-921. <https://doi.org/10.2307/376528>
- Fischer-Starcke, B. (2009). Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A corpus-stylistic analysis. *International*

- Journal of Corpus Linguistics*, 14(2), 492–523. <https://doi.org/10.1075/ijcl.14.4.03fis>
- Fischer-Stracke, B. (2010). *Corpus linguistics in literary analysis: Jane Austen and her contemporaries*. Continuum.
- Flynn, E. (1988). Composing as a woman. *College Composition and Communication*, 39(2), 423-435. <https://doi.org/10.2307/357697>
- Furumo, K., & Pearson, M. (2007). Gender-based communication styles, trust, and satisfaction in virtual teams. *Journal of Information, Information Technology & Organizations*, 2(2), 49-60. <https://doi.org/10.28945/138>
- Gilligan, C. (1982). *In a different voice*. Harvard University Press.
- Granger, S. (1996). From CA to CIA and back: An integrated approach to computerised bilingual and learner corpora. In K. Aijmer (Ed.), *Languages in contrast. Text-based cross-linguistic studies* (pp. 37-51). Lund University Press.
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7-24. <https://doi.org/10.1075/ijlcr.1.1.01gra>
- Hiatt, M. (1977). *The way women write*. Teachers College Press.
- Hiatt, P. (1976). The sexology of style. *Language and Style*, 9(2), 98-107.
- Holmes, D. I. (1998). The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing*, 13(3), 111-117. <https://doi.org/10.1093/lc/13.3.111>
- Hunter, P., Pearce, N., Lee, S., Goldsmith, S., Feldman, P., & Weaver, H. (1988). Competing epistemologies and female basic writers. *Journal of Basic Writing*, 7(1), 73-81. <https://www.jstor.org/stable/43443496?seq=1>
- Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104(1), 53-69. <https://doi.apa.org/doi/10.1037/0033-2909.104.1.53>
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192. <https://doi.org/10.1177/1461445605050365>
- Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6(2), 183-205. [https://doi.org/10.1016/S1060-3743\(97\)90033-3](https://doi.org/10.1016/S1060-3743(97)90033-3)
- Ishikawa, S. (2011). A new horizon in learner corpus studies: The aim of the ICNALE project. In G. Weir, S. Ishikawa, & K. Poonpon (Eds.), *Corpora and language technologies in teaching, learning and research* (pp. 3-11). University of Strathclyde Press.
- Johnston, E., Berry, J., & Mielke, W. (2006). Measures of effect size for chi-squared and likelihood-ratio goodness-of-fit tests. *Perceptual and Motor Skills*, 103(2), 412-414. <https://doi.org/10.2466/2Fpms.103.2.412-414>
- Jones, S., & Myhill, D. (2007). Discourses of difference? Examining gender differences in linguistic characteristics of writing. *Canadian Journal of Education*, 30(2), 456-482. <https://doi.org/10.2307/20466646>
- Jule, A. (2017). *A beginner's guide to language and gender*. Multilingual Matters Ltd.
- Koppel, M., Argamon, S., & Shmoini, A. R. (2002). Automatically categorising written texts by author gender. *Literary and Linguistic Computing*, 17(4), 401-412. <https://doi.org/10.1093/lc/17.4.401>
- Krijnen, T., & van Bauwel, S. (2015). *Gender and media: Representing, producing, consuming*. Routledge.
- Kunter, G. (2017). CoQuery (Version 0.10.0) [Software]. Available from <https://www.coquery.org/>
- Lakoff, R. (1972). Language in context. *Language*, 48(2), 907-924. <https://doi.org/10.2307/411994>
- Lakoff, R. (1973). Language and woman's place. *Language in Society*, 2(1), 45-79. <https://doi.org/10.1017/S0047404500000051>
- Leeper, C., & Ayres, M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. *Personality and Social Psychology Review*, 11(4), 328-363. <https://doi.org/10.1177/1088868307302221>
- Maltz, N., & Borker, A. (1982). A cultural approach to male-female miscommunication: A cultural approach to interpersonal communication. *Essential Readings*, 16(2), 8-185. <https://doi.org/10.1017/cbo9780511620836.013>
- Mulac, A., & Lundell, L. (1994). Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language effects. *Language & Communication*, 14(3), 299-309. [https://doi.org/10.1016/0271-5309\(94\)90007-8](https://doi.org/10.1016/0271-5309(94)90007-8)
- Mulac, A., Lundell, L., & Bradac, J. (1986). Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the gender-linked language effect. *Communications Monographs*, 53(2), 115-129. <https://doi.org/10.1080/03637758609376131>
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher* 31(7), 9-13. <https://jalt->

- publications.org/sites/default/files/pdf/the_language_teacher/07_2007tlt.pdf
- Oakley, A. (1972). *Sex, gender and society*. Temple Smith.
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the workshop on comparing corpora-Volume 9* (pp. 1-6). Association for Computational Linguistics.
- Roen, D. H., & Johnson, D. M. (1992). Perceiving the effectiveness of written discourse through gender lenses. The contribution of complimenting. *Written Communication*, 9(4), 435-464.
- Rubin, L., & Greene, K. (1992). Gender-typical style in written language. *Research in the Teaching of English*, 26(1), 7-40. https://www.researchgate.net/publication/292650726_Gender-Typical_Style_in_Written_Language
- Rubin, L., & Nelson, M. W. (1983). Multiple determinants of a stigmatised speech style: Women's language, powerless language, or everyone's language? *Language and Speech*, 26(3), 273-290. <https://doi.org/10.1177%2F002383098302600308>
- Samar, R., & Shirazizadeh, M. (2010). Gender-preferential linguistic elements in applied linguistics research papers: Partial evaluation of a model of gendered language. *Applied Linguistics*, 3(1), 70-84. http://jal.iaut.ac.ir/article_523287.html
- Sarawgi, R., Gajulapalli, K., & Choi, Y. (2011). Gender attribution: Tracing stylometric evidence beyond topic and genre. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning* (pp. 78-86).
- Scates, C. (1981). *A sociolinguistics study of male/female in freshman composition*. University of Southern Mississippi Press.
- Schirmer, A., Kotz, A., & Friederici, D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research*, 24(3), 442-452. <https://doi.org/10.1016/j.cogbrainres.2005.02.022>
- Smeltzer, L. R., & Werbel, J. D. (1986). Gender differences in managerial communication: Fact or folk-linguistics? *The Journal of Business Communication* (1973), 23(2), 41-50. <https://doi.org/10.1177%2F002194368602300205>
- Snape, N. (2008). *The acquisition of the English determiner phrase by L2 learners: Japanese and Spanish*. VDM Verlag.
- Sterkel, S. (1988). The relationship between gender and writing style in business communications. *The Journal of Business Communication*, 25(4), 17-38. <https://doi.org/10.1177%2F002194368802500402>
- Stubbs, M. (2005). Conrad in the computer: Examples of quantitative stylistic methods. *Language and Literature*, 14(1), 5-24. <https://doi.org/10.1177%2F0963947005048873>
- Sunderland, J. (2000). New understandings of gender and language classroom research: Texts, teacher talk, and student talk. *Language Teaching Research*, 4(2), 149-173. <https://doi.org/10.1177%2F136216880000400204>
- Swacker, M. (1975). The sex of the speaker as a sociolinguistic variable. In B. Thorne & N. Henley (Eds.), *Language and sex: Difference and dominance* (pp. 76-83). Newbury House Publishers.
- Swales, J. (1990). *Genre analysis*. Cambridge University Press.
- Talbot, M. (1998). *Language and gender*. Polity.
- Tannen, D. (1990). *You just don't understand: Men and women in conversation*. Morrow.
- Taylor, S. (1978). Women in a double bind: hazards of the argumentative edge. *College Composition and Communication*, 29(4), 385-389. <https://doi.org/10.2307/357026>
- Zheng, R., Li, J., Chen, H., & Huang, Z. (2006). A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the Association for Information Science and Technology*, 57(3), 378-393. <https://doi.org/10.1002/asi.20316>