

Tanulmány

Xiaoyun Li

Involvement in Learner English Writing: The Case from Asian Learners

Abstract

Involvement refers to the engagement of the speaker or hearer in discourse. As the most typical feature of spoken language, it has been regarded as one of the major features of L2 writing due to writers' lack of register awareness. However, a closer look at a number of related studies reveals that the investigated language data are not well selected and that Asian learners of English are less targeted, which might make their research findings less convincing. By employing the 'Involved versus informational production' dimension from Biber's 1988 multidimensional analysis (MDA) model, this study sets out to verify the previous findings regarding involvement based on a well-balanced Asian learner English corpus – International Corpus Network of Asian Learners of English (ICNALE). The results reveal that learner English writings are not necessarily more involved than the English writings by native English speakers as English learners from Asian ESL contexts including Hong Kong, Philippines, Pakistan, and Singapore, show an informational tendency; learners' L1 affects their usage of some specific involvement features but it, in general, does not affect the overall involvement in their writings; learners' proficiency and involvement are in a negative correlation; female Asian learners generally tend to be less involved than their male peers in writing, which is contrary to the case of native English speaker group; gender differences regarding involvement are more significant in L1 writing than in L2 writing;

Keywords: Involvement, Learner English Writing, Asian learners of English, Multidimensional analysis.

1 Introduction to Involvement

Involvement is a hot concept in the discussion on the differences between spoken and written language. On a basis of massive observation of the features of the two languages, researchers (Chafe 1985; Biber 1988; Biber et al. 1999) point out that spoken language commonly contains a great number of features that mark the involvement of the speaker or hearer, while written language, on the other hand, tends to be more informational and less involved owing to the frequent use of informational and detachment features (features that help establish the detachment of the speaker or hearer), including nouns, adjectives, impersonal pronouns, etc. For the different involvement degrees of the two forms of language, researchers have offered different explanations from varied perspectives. The most basic is that the relation between the language producer and receiver varies in the spoken and written context. In spoken language, target receivers are physically present, and immediate feedback between speakers and hearers is available, but in written language readers are often unseen or unknown. Therefore, Chafe & Danielwicz (1987) conclude that "spoken language contains indications of the speaker's

Xiaoyun Li:

Involvement in Learner English writing: The Case from Asian Learners

Argumentum 18 (2022), 124–143

Debreceni Egyetemi Kiadó

DOI: 10.34103/ARGUMENTUM/2022/7

involvement with the audience, with himself/herself and with the concrete reality whilst written language lacks those kinds of involvements” (Chafe & Danielwicz 1987: 19).

Tannen (1983) interprets involvement as an oral strategy through which language users can achieve their communicative goals. In talking about the comparison between spoken and written language, she states that the differences between the two forms of language are more likely to originate from language users' relative focus on interpersonal involvement and content-conveying, and that oral strategies (associated with interpersonal involvement) may contribute “successful discourse production and comprehension in the written as well as the oral mode” (Tannen 1983: 92). Similarly, findings from Hubbard's research (2001) confirm that the increase in involvement is found to be connected with higher readability of texts and that involvement strategies associated with spoken language might be the most efficient for writing and reading. Biber (1988) observes involvement from a functional perspective and interprets involvement as an umbrella term referring to the features marking affective, interactional, and generalized content, and these features, in his multi-dimensional model accounting for the variation across written and spoken registers, functionally cluster in most of the spoken registers to perform an involved focus.

Due to different interpretations of involvement, the involvement features adopted by researchers vary. The most accepted involvement features are outlined by Chafe (1985). According to the objects being involved, Chafe distinguishes three types of involvement with each type associated with different sets of linguistic devices – ego involvement (e.g., first-person pronouns), involvement with the audience/hearer/reader (e.g., second-person pronouns and questions), and involvement with the subject matter (e.g., direct quotations). This categorization is followed by Altenberg (1997), Ädel (2008), Bednarek (2014), and partially by Petch-Tyson (1998). Involvement features utilized in Barbieri's (2013) research on analyzing university classroom discourse are the most detailed – in total 18 involvement markers are summarized through surveying relevant literature and consulting Biber et al. (1999)'s Longman Grammar.

2 Involvement in Learner Language

Involvement has been regarded as the most salient feature of learner English productions in many related studies. Studies conducted by (Ädel 2008), Aijmer (2001), Gilquin & Paquot (2008), Lee et al. (2019), Petch-Tyson (1998), Recski (2004), Yoon (2015), etc., reveal that in comparison with their native counterparts, L2 learners, irrespective of their L1s, exhibit an overuse of linguistic features which are regarded as informal and closely associated with spoken language, such as first-person pronouns, private verbs, second-person pronouns, direct questions, and emphasizees. These studies consequently suggest that the high involvement found in L2 writings demonstrates L2 learners' poor register awareness and emphasize the necessity of raising learners' register awareness in language teaching and writing.

To account for the high involvement in learner writings, several explanations have been proposed. In an investigation of the writings composed by L2 English learners with different L1 backgrounds from Europe, Petch-Tyson (1998) finds that different groups of L2 learners reveal quantitative and qualitative differences in the use of 1st person pronoun 'I' and argues that it might be “culturally induced” (Petch-Tyson 1998: 117). Similarly, Kobayashi & Abe (2016) substantiate the influence of learners' L1s on their register awareness by comparing four

Xiaoyun Li:

Involvement in Learner English writing: The Case from Asian Learners

Argumentum 18 (2022), 124–143

Debreceni Egyetemi Kiadó

DOI: 10.34103/ARGUMENTUM/2022/7

sub-corpora from the International Corpus Network of Asian Learners of English (ICNALE). They note that learner writings from Hong Kong display a set of stylistically appropriate features, whereas those from Japan, Korea, and Taiwan show an informal and spoken-like tendency. Friginal et al. (2014) confirm that high-rated essays are in a positive relationship with the informational writing style, i.e., students who wrote in a more informational style tended to get higher scores than students who wrote in a more personal style. Gilquin & Paquot (2008) explore the spoken features in learner academic writings contained in the International Corpus of Learner English (ICLE) and summarize four factors that might be accountable for the spoken-like nature of learner writing, namely, the influence of speech, L2 instruction, developmental factors, and L1 transfer. Crosthwaite (2016) and Issitt (2016) apply multi-dimensional analysis (MDA) to examine the effectiveness of EAP courses on writing and find a positive effect of L2 instruction in diminishing oral features. Situational factors are also proved to be related to involvement. Ädel (2008) investigates the involvement features in English essays by Swedish speakers and argues that time availability (timed versus un-timed) and intertextuality (access to secondary sources) affect involvement in learner writings.

Despite the extensive literature on involvement in learner English, some issues remain still. First of all, few studies have targeted Asian learners. As the above literature survey shows, the bulk of involvement study centers on learners of English from Europe. The present study, therefore, hopes to contribute to this under-researched area. Secondly, a fair amount of research focuses on Asian learners from mono-mother-tongue background. Even those studies which claim to be Asian learner-centered are limited in the amount of L2 learner groups, which might limit our understanding of involvement. Thirdly, the corpora investigated in some related studies are not comparable enough though the research results are illuminating and insightful. In some studies, the compared corpora are different in terms of topics or genres, for example, in Gilquin & Paquot's two studies (2007 & 2008), the written BNC written corpus which consists of samples from books and journals in different disciplines is compared to ICLE which comprises 700-words-long essays that are composed in different task settings. Yet, this is not to say that corpus research can only be done on corpora that perfectly match. Instead, what is stressed here is that it might be more plausible if the corpora being used are more comparable because it is highly possible for variables including genres, task settings, topics, etc., to skew the authentic research findings. As one of the largest Asian L2 composition databases, ICNALE provides a representative and well-balanced language database to explore involvement in Asian learner English.

3 The Present Study

Motivated by the issues outlined above, the present study intends to analyze involvement in the English writings by learners of English from different countries and regions of Asia under the guidance of the following three hypotheses that are summarized based on related literature:

- 1) Compared with English writings by native speakers, Asian learner writings might be typical of involvement, given abundant research suggests that native English writing contains fewer involvement features (Cobb 2003; Paquot et al. 2013).
- 2) Learners' L1s and proficiency levels might affect involvement in their English writings, given most of the related research suggests that the two factors lead to the variation of

Xiaoyun Li:

Involvement in Learner English writing: The Case from Asian Learners

Argumentum 18 (2022), 124–143

Debreceni Egyetemi Kiadó

DOI: 10.34103/ARGUMENTUM/2022/7

involvement in learner language (Gulquin & Paquot 2007 & 2008; Kobayashi & Abe 2016).

- 3) Female learners of English might show higher involvement than their male peers in writing, given that in many studies, female language is believed to be more interactional while male language is more informational (Holmes 1995; Lakoff 1973; Argamon et al. 2003).

Unlike the foregoing studies, the present study delves into involvement in Asian learner English writings by adopting the linguistic features included in the first dimension of Biber's 1988 multidimensional analysis (MDA) model (1988) – “involved versus informational production”. This decision is made out of two considerations. The first is that this dimension is the biggest factor that explains the variation between spoken and written registers. Among the 29 investigated features in dimension 1, 23 features that are identified as typical in spoken registers largely cover the involvement features investigated in related literature. Besides, although the extra involvement features listed in Barieri (2013)'s study might provide a good supplement, the present study does not include these features out of the consideration that they are highly oral and rarely appear in written texts, for example, hesitators like “*mhm/uh*”. The second is that the first dimension in Biber's model is a continuum ranging from the pole of informational to involved production and the features investigated also include those that are identified as informational. This is in accordance with a widely recognized view that differences between spoken and written language should be conceived of as a continuum instead of a dichotomy (Ädel 2008; Tannen 1982) because their typical linguistic features are not exclusive in essence (Lingley 2005). Most of the previous literature, however, focuses entirely on “spoken features (or involvement features)”, but ignores the fact that informational features are crucial as well for observing the difference between the two forms of language on involvement, thus unintentionally following the dichotomy theory. As will be seen in the following, our understanding of involvement in L2 writings might be badly distorted if no heed is paid to informational features.

4 Data and Method

The learner data analyzed in this study were extracted from the Written Essay Module of ICNALE built by Ishikawa (2013 & 2014). ICNALE is the largest Asian learner English corpus ever compiled (Ishikawa 2013), with 4 modules (Spoken Monologues, Spoken Dialogues, Written Essays, and Edited Essays) that consist of 3.55 million (by 2021 September) words of language data produced by L2 learners of English from 10 different Asian countries and regions as well as by native English speakers from several major English-speaking counties. This corpus is well designed in that the contained data are strictly controlled concerning registers, topics, composing time, etc., making it a reliable data source for varied types of contrastive interlanguage analysis (Ishikawa 2013). In the Written Essay Module, 2800 participants, including both Asian learners of English (A2 to C1 learners) and native English speakers (novice and experienced L1 writers), were asked to write two short English essays (200 to 300 words) on two argumentative topics (*It is important for college students to have a part-time job* and *Smoking should be completely banned at all the restaurants in the country*), within 80 minutes. Currently, it contains 5600 English writings, totaling approximately 1.3 million words.

In this study, all the 11 written corpora (10 Asian learner corpora and 1 native English corpus) of the Written Essay Module were selected (See Table 1).

As mentioned earlier that the features investigated in the present study are taken from the first dimension of Biber’s 1988 MDA model (Biber 1988), it is necessary to briefly introduce Biber’s 1988 MDA model before probing into involvement in Asian learner English writings. This model originates from Biber’s research on the variation across spoken and written registers in English (Biber 1988). With the assumption that co-occurring linguistic features reflect shared functions in mind, Biber conducted a factor analysis to identify the co-occurrence pattern of linguistic features in Lancaster-Oslo-Bergen Corpus of British English and London-Lund Corpus of Spoken English. By interpreting the shared functions of the clustered linguistic features in each retained major factor, Biber obtained a 6-dimension MDA model in which each dimension “defines the continuum of variation rather than discrete poles” (Biber 1988: 9). To take dimension 1, which is adopted in this study, as an example, it is a continuum ranging from the pole of informational production to the pole of involved production. Oral or information texts would be in a higher position as they make frequent use of involvement features (features with positive loadings) but lack informational features (features with negative loadings), while formal written texts are typically high in the use of informational features but low in involvement features and thus are normally located at a lower position of the continuum.

Corpus	Number of texts	Number of words
Mainland China (CHN)	800	200,516
Philippine (PHL)	400	98,550
Native English speaker (NES)	400	89,513
Korea (KOR)	600	135,264
Japan (JPN)	800	177,285
Taiwan (TWN)	400	91,420
Hong Kong (HKG)	200	47,358
Thailand (THA)	800	179,967
Singapore (SIN)	400	98,978
Indonesia (IDN)	400	92,867
Pakistan (PAK)	400	94,361
Total	5600	1,306,079

Table 1. Written corpora selected from ICNALE

To conduct a multidimensional analysis is quite demanding and time-consuming as it not only involves MDA POS-tagging but also requires a large amount of computation. Luckily, the appearance of MAT (Multidimensional Tagger), an automatic tool which is specially designed by Nini (2012) for conducting multidimensional analysis within Biber’s 1988 MDA model, significantly reduces the workload. The functions of MAT generally can be divided into two categories: MDA POS-tagging and dimension score calculating. The MDA POS-tagging function of it is achieved via using 67 linguistic features of Biber (1988). The dimension score calculating function of MAT encompasses two procedures. The first one is to obtain the standardized scores of the linguistic features based on the nominalized mean and standard deviation values offered in Biber (1988). The computing formula is $z\text{-score} = \frac{\chi - \mu_B}{\sigma_B}$ (Nini 2019), in which χ , μ_B , and σ_B represent respectively *the normalized frequency of*

the feature in consideration, Biber's mean normalized frequency of the feature in consideration, and Biber's standard deviation of the feature in consideration. The second procedure is to calculate the dimension scores by summing or subtracting the z-scores of the linguistic features. To take Dimension 1 as an example again, the computing formula is:

$$\text{Score on Dimension 1} = (Z_{\text{privateverb}} + Z_{\text{1st personpronoun}} + \dots + Z_{\text{contraction}}) - (Z_{\text{noun}} + Z_{\text{adjective}} + \dots + Z_{\text{preposition}}) \text{ (Nini 2019).}$$

The z-scores of positive loading features on Dimension 1 are summed firstly, then the received sum subtracts the sum of the z-scores of negative loading features to obtain the dimension score.

Once a text is entered, MAT will automatically conduct MDA tagging and dimension score calculating. The dimension score of a corpus is the mean of the dimension scores of the texts it contains. Besides dimension scores, the returned results of MAT also include the frequency information of each linguistic feature, which provides great convenience for doing a deeper investigation given that corpora with similar dimension scores might differ with regard to their usage of linguistic features and the frequency information can make a good supplement to the observation of the differences between corpora. In this study, the newest version of MAT, MAT 1.3, was adopted (its list of the variables on Dimension 1 is presented in Table 2).

Features with positive loadings	Private verbs (PRIV), subordinator that deletion (THATD), contractions (CONT), present tense verb (VPRT), 2nd person pronouns (SPP2), do as pro-verb (PROD), analytic negation (XX0), demonstrative pronouns (DEMP), emphatics (EMPH), 1st person pronouns (FPP1), the pronoun <i>it</i> (PIT), <i>be</i> as main verb (BEMA), causative adverbial subordinator (CAUS), discourse particles (DPAR), indefinite pronouns (INPR), hedges (HDG), amplifiers (AMP), sentence relatives (SERE), direct <i>wh</i> -questions (WHQU), possibility modals (POMD), independent clause coordination (ANDC), <i>wh</i> -clauses (WHCL), stranded preposition (STPR).
Features with negative loadings	Average word length (AWL), type-token ratio (TTR), attributive adjectives (JJ), nouns (NN), prepositions (PIN).

Table 2. Linguistic features of Dimension 1 in Biber's (1988) dimension model

5 Results and Discussion

Hypothesis 1: Compared with the writings by native English speakers, Asian learner writings might be typical of involvement.

From Figure 1 where the scores of the 11 corpora are presented, we can see that it would be problematic to conclude that Asian learner writings are typical of involvement as compared with native English writings: the native corpus is indeed lower than those from the EFL context (i.e., TWN, THA, JPN, KOR, CHN, IDN) in terms of their dimension scores whereas it is higher than the learner corpora from ESL contexts (PHL, PAK, HKG, SIN). More specifically, learner writings from the EFL contexts do exhibit an overtly involved style while those from ESL contexts show lower involvement than native English writings. To draw a more valid conclusion, this study conducted *t*-tests between the learner corpora and the native corpus, and the results

are also presented in Figure 1. It can be seen that all the corpora from EFL contexts except for IDN (but it shows no statistically significant difference from CHN), are statistically different from the native speaker corpus. Similar findings can be observed on the corpora from ESL contexts as well, with the exception of PHL. Among the corpora from ESL contexts, PHL appears to be an outlier. It resembles NES but differs considerably from the other ESL learner corpora, namely, PAK, HKG, and SIN. The explanation might be related to the differences between American English (AmE) and British English (BrE) as PHL differs from the other ESL learner corpora in that it is from an ESL context dominated by AmE (Martin 2014). Besides, the fact that the majority of writers of NES are from the U.S., along with the finding of Biber (1987) that American written genres are more colloquial and interactive than the corresponding British genres, also adds extra credibility to the explanation.

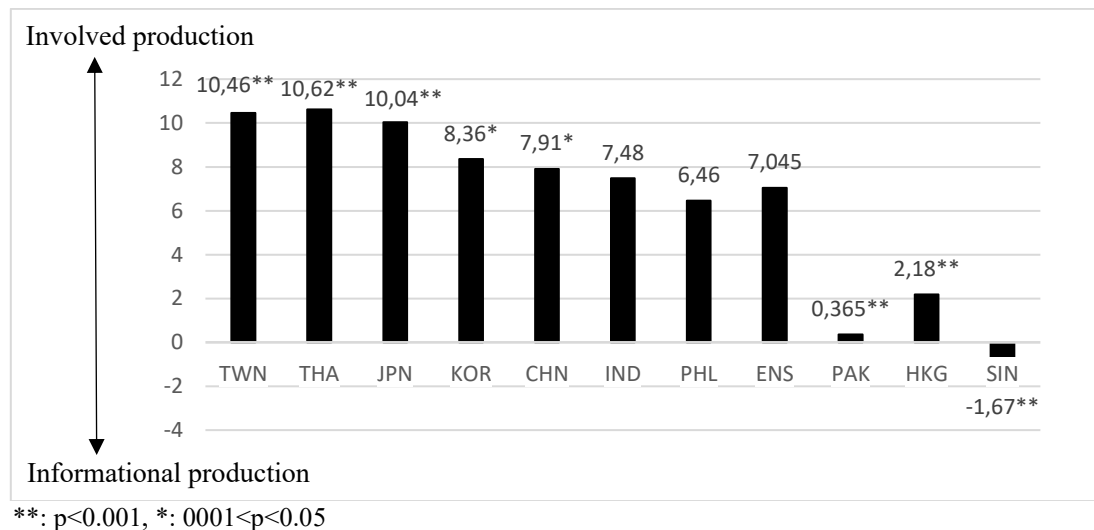


Figure 1. Dimension scores of the 11 corpora

Appendix I reveals that EFL learners, ESL learners, and native English speakers show different preferences on the usage of linguistics features included in dimension 1. Writings by ESL learners, in comparison with those by ESL learners and native English speakers, are high on the use of the features with positive loadings, especially ANDC, CONT, FPP1, SPP2, PRIV, VPRT, and XX0, whereas they obtain relatively small values on negative-loading features, indicating that they are informational and at the same time, highly involved in terms of writing style. Writings by ESL learners show less use of a series of major positive features, including ANDC, CONT, FPP1, EMPH, PRIV, PROD, SPP2, THATD, and VPRT, but surpass the writings by EFL learners and native English speakers on AWL, NN and PIN, in particular, suggesting a non-involved and informational concern. As for native English speakers (NES), their writings make relatively more use of positive features than ESL learner writings but less than EFL learner writings. On the usage of negative features, writings by native speakers are close to those by ESL learners.

Seen from a perspective of language exposure, the above findings might not be surprising. Alhusban & Vijayakumar (2021) point out that “our use of lexical resources is largely determined by our prior knowledge of the contexts (registers and genres) and the exposure we received” (Alhusban & Vijayakumar 2021: 23). Inexperienced English L1 writers who do not

Xiaoyun Li:

Involvement in Learner English writing: The Case from Asian Learners

Argumentum 18 (2022), 124–143

Debreceni Egyetemi Kiadó

DOI: 10.34103/ARGUMENTUM/2022/7

receive systematic writing instruction and reading training, are more likely to resort to spoken language instead of written language in writing due to extensive contact with spoken language. Comparatively, ESL learners are exposed less to spoken English, but more to formal written English as English, more often than not, is utilized in formal domains of ESL societies, for example, education, law, business, etc. Moreover, the English acquisition or learning of L2 learners is primarily conducted in the classroom context through the way of teacher instruction, and they might be “socialized in educational contexts in which academic lectures are modeled based on formal written texts” (Barbieri 2013: 170). In accounting for the low percentage of ego-involvement in ESL student writing, Hyland (2002a & 2002b) and Lee et al. (2019) agree that the recommendations from writing guides and teachers' instruction are the two major factors leading to ESL learners' strict adherence to an impersonal and low author-involved writing style.

In a similar vein, experienced English L1 writers who receive substantial exposure to English academic literacy have a better grasp of written register knowledge. With rich writing experience and relevant register knowledge, they are more capable of exploiting involvement features to achieve their writing purposes without concerning the register appropriateness. Lee et al. (2018) compare top-rated essays written by English L1 and L2 writers and find that English L1 writers are liberal in adopting traditional prescriptive writing conventions and informal language devices such as first-person pronouns/determiners, and contractions. Additionally, for experienced English L1 and L2 writers, involvement, particularly ego-involvement, performs special rhetorical functions. In a fundamental sense, involvement in writing refers to the participation or engagement of the writer or the reader. As a result, involvement features, especially those which are believed to be closely linked to subjectivity and personal attachment, such as *first-/second-person pronouns* and *private verbs*, perform important discourse functions including marking authority (viz, claiming the author's responsibilities for the content or contribution), highlighting independent voice, guiding readers, etc. (Harwood 2005; Hyland 2002a & 2002b).

Then, why does the previous research conclude that learner writings are higher in involvement than native English speakers' writings? One obvious explanation is that they did not survey enough L2 learner groups, especially those from ESL contexts in Asia. Given that the findings are based on language data from a limited number of L2 learner groups, the generalization of their research findings is limited accordingly. Similarly, the above findings need to be treated with caution as well since other Asian ESL contexts like India and Malaysia are left unexplored. Another explanation might be that the corpora being compared in previous studies are not comparable enough. For instance, the comparisons between the International Corpus of Learner English (ICLE) and the academic component of the British National Corpus (BNC) made in Gilquin & Paquot's (2007 & 2008) exploration into spoken features in learner academic writing is debatable. Although the language data the two researchers extracted are all called academic, they lack comparability: the academic component of BNC (15 million words) is comprised of samples from academic books, published and unpublished letters and memoranda, school and university essays from various academic fields, whereas ICLE consists of around 6000 L2 English argumentative essays totaling approximately 3.5 million words (Gilquin & Paquot 2007). In other words, the criteria employed in the two corpora for text sampling vary markedly. Lastly, researchers might not distinguish between writings by experienced and by novice writers on choosing a reference corpus of native English for comparison. The pronounced difference between the two NES corpora on involvement shown

in Figure 2 reflects the high heterogeneity of the Native English speaker group. Undoubtedly, comparing writings by experienced or professional English L1 writers to those by L2 learners of English is of great pedagogical implications (Ackerley 2008). Nevertheless, it cannot lead us to the conclusion that L2 learner writing is typical of involvement.

Hypothesis 2: learners' L1s and proficiency levels might affect involvement in their English writings.

The first aspect of this hypothesis focuses on the interference of learners' mother tongue in their written English productions on involvement. Figure 1 shows that learners' mother-tongue background seems not to be a factor causing the variation of involvement in their writings: TWN, THA and JPN receive similar scores though their L1s are different, and the same finding applies to the rest of EFL corpora (i.e., KOR, CHN, and IDN) and to the four ESL corpora, PHL, PAK, HKG, and SIN. More importantly, HKG closely resembles PAK and SIN, instead of TWN and CHN – two corpora that are inextricably linked with HKG concerning L1 and culture, which strongly disproves the hypothesis. The frequency information of HKG in Appendix I also reveals that HKG is quite distinct from CHN and TWN though they share the same mother tongue. Furthermore, the distribution of the corpora in Figure 1 displays a division between ESL and EFL groups if NES is taken as the benchmark, suggesting that the English learning context plays a more important role than learners' L1 in altering involvement in their writings.

The above finding, however, does not imply that learners' L1 is irrelevant to involvement in their writings. Considering that L1 transfer is widely acknowledged (Hu & Bodomo 2009) and that the dimension scores are calculated by summing up the loading weights of each feature, this study inclines to the view that the transfer of L1 is likely exhibited in learners' usage of specific involvement features. In fact, relating involvement features that frequently appear in both L2 learner target language production to L1 transfer has been a popular way in related literature. In the two studies that form the basis of the hypothesis, namely, Gulquin & Paquot (2007 & 2008), the frequent use of “*let's/let us*” of French-speaking learners of English was cited as a piece of L1 transfer evidence to explain the spoken style in learners' writings. Scrutiny of the normalized feature frequencies from TWN and CHN (see Appendix I) shows that although the two corpora obtain different dimension scores, they share a similar pattern on the usage of the investigated features except for minor differences on the frequencies of XX0, VPRT, SPP2, and PIN. Another supporting argument is the abnormally high frequency of FPP1 in JPN. In this study, all the four proficiency levels of Japanese Learners of English are found to be overusing *I*: 11.41 *I* per essay for A2 learners, 11.45 for B1_1, 10.28 for B1_2, and 11.11 for B2. In his investigation of *I* in the English essays by Japanese EFL learners, Natsukari (2012) ascribes the frequent appearance of *I* to the Japanese language because “Japanese discourse is always expressed from a speaker's point of view ...” (Natsukari 2012: 72) and therefore, “typical and traditional composition written in Japanese is full of subjectivity” (Natsukari 2012: 73).

As for the strikingly low involvement of HKG in comparison with CHN and TWN, the possible reasons might lie in the special ESL context of Hong Kong and the relatively high English proficiency of Hong Kong learners. Due to the ESL context, learners from Hong Kong receive more exposure to English than their counterparts from Mainland China and Taiwan, and the exposure is more likely about formal written English instead of spoken English as spoken English in Hong Kong has diminished while written English has remained “firmly entrenched”

(Evans 2011: 302) in domains including education, government, business and law since 1997. With regard to the latter reason, learners in HKG are above B1 level while learners in CHN and TWN range between A2 to B2 as far as their English proficiency is concerned, and as we will see in the following discussion, learners' proficiency cannot be ignored in accounting for the involvement variation of their writings.

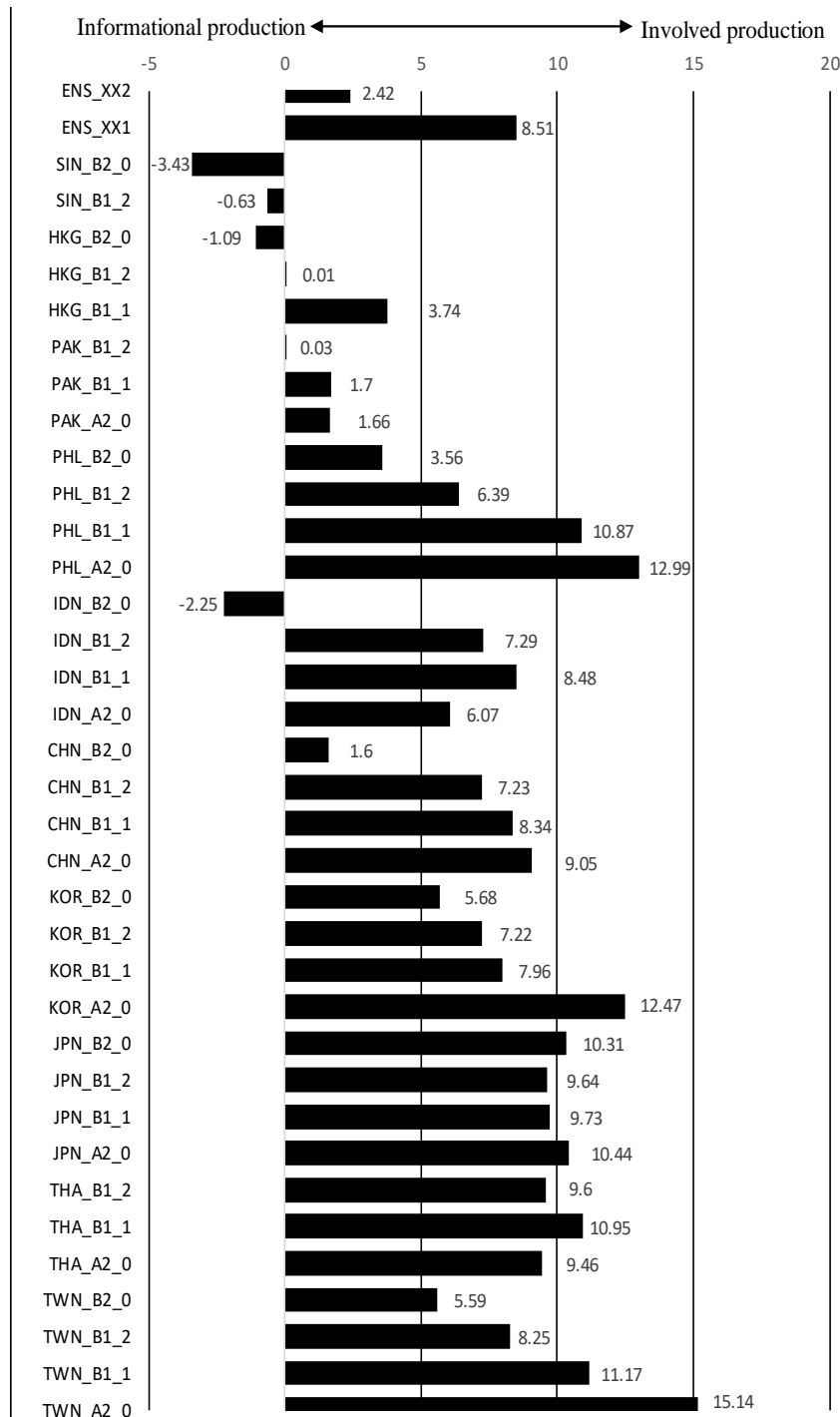


Figure 2. Language proficiency and dimension scores

Xiaoyun Li:

Involvement in Learner English writing: The Case from Asian Learners

Argumentum 18 (2022), 124–143

Debreceni Egyetemi Kiadó

DOI: 10.34103/ARGUMENTUM/2022/7

The second aspect of Hypothesis 2 concerns the link between learners' target language proficiency and involvement in their writings. In Figure 2, eight out of the ten investigated L2 English corpora show that involvement degree decreases as learners' proficiency advances, generally confirming the effect of learners' proficiency on involvement in their writings. A close examination of the detailed usage of the linguistic features shows that this negative correlation relationship between involvement and language proficiency is manifested by the increase of negative loading features while the reduction of positive loading features. More specifically, writings by advanced learners are characterized by high frequencies of informational features but low frequencies of involvement features, and vice versa for those by less advanced learners.

The high number (or value) of informational features found in advanced learner writing can be attributed to their superiority in linguistic resources since all the features are directly associated with learners' master of English vocabulary. Grant & Ginther (2000) points out that AWL and TTR “are indicative of sophisticated writing” (Grant & Ginther 2000: 130) and higher proficiency writers are “more precise about using words that best express their ideas” (Grant & Ginther 2000: 131). It is reasonable to expect that compared with low proficiency learners, advanced learners have fewer problems in the aspects of word-finding, grammar, etc., due to their extensive linguistic choices. Their linguistic superiority puts them in a better position to compose a sophisticated essay by using sophisticated words that are commonly lengthy (Wolfe-Quintero et al. 1998). By the same token, advanced learners are more skillful in packing plenty of information into an essay by massively exploiting NN, JJ, and PIN, making the contained arguments more informational.

Advanced learners' superiority in linguistic resources can also contribute to the lowering of the frequencies of involvement features in their writings. Sufficient linguistic choices empower advanced learners with more alternative expressions to replace involvement features which, according to Ädel (2008), are learned by L2 learners at the beginning stages of language learning, especially some common but simple lexical items including personal pronouns, amplifier *very*, etc. What makes the need of advanced L2 learners for making the substitution more necessary is that low lexical repetition is one of the basic features of good writing (Read 2000). It therefore can be expected that some common involvement features used by the less advanced learners in writing would be intentionally avoided by advanced learners to receive good assessments. To take the collocation of 1st person-pronoun *I* and the private verb *think* as an example, it is found in the present study that less proficient learners have a tendency to overuse this collocation in marking authority and personal stance while in advanced learner writings, it is often replaced with phrases with similar structures such as *I agree/believe/disapprove*, passive voice structures starting with anticipatory *it, as far as I am concerned*, to name but a few.

Hypothesis 3: Female learners of English might show higher involvement than their male peers in writing.

This hypothesis is proposed with an attempt to verify whether involvement in learner writings is subject to the influence of the author's gender. According to the gender information offered by ICNALE, every corpus investigated in the study was further divided into the female and male corpora. The corpora were then compared in terms of their dimension scores. Moreover, to better confirm the possible differences between the two genders, the independent t-test was adopted.

Regions	Female (mean)	Male (mean)	<i>t</i> -value	<i>p</i> -value
PHL	6.73	6.28	0.440	0.730
PAK	-0.36	1.09	-1.350	0.179
NES	9.43	4.66	4.257	<0.001
KOR	7.46	9.25	-1.583	0.114
JPN	10.15	9.97	0.273	0.785
TWN	9.99	10.93	-0.650	0.517
THA	10.47	10.77	-0.365	0.716
HKG	1.26	3.1	-1.152	0.252
SIN	-2.15	-1.18	-1.166	0.245
IDN	6.11	8.84	-2.603	0.015
CHN	7.22	8.6	-1.334	0.220

Table 3. Comparison between the male and female writings

It can be seen from Table 3 that the two genders of the learner corpora score rather closely, and the results of *t*-tests also confirm that nine out of the ten investigated learner corpora do not show any statistically significant differences between the two genders on involvement. Nevertheless, no statistically significant differences do not imply that the two genders are the same on involvement. A closer look at Table 3 indicates that it seems plausible to suppose that female learner writings, contrary to the hypothesis, are relatively low on involvement in comparison to male learner writings. Not to mention IDN in which female learner corpus is much lower than its male counterpart on dimension score and the difference between them is shown to be statistically significant, there are eight learner corpora in which the female corpora score lower than the male corpora. Even in the two exceptions, PHL and JPN, the scores the female corpora obtained are only marginally higher than what their male counterparts receive. The difference between the two genders on involvement is also manifested by their inconsistency over the usage of the linguistic features of dimension 1, especially, the positive features. In Appendix II where the two genders' detailed usage of linguistic features is presented, the male learner corpora appear to be containing a slightly higher number of positive features.

In the discussion of the previous two hypotheses, written language exposure and language proficiency are considered to be affecting involvement in L2 learner writing. It is hard for this study to ascertain whether there is a difference in exposure to written language between the two genders since this topic is beyond the scope of the present study. However, there is a large body of research proving that females are better learners in L2 language learning compared to males (Główka 2014; Bernhard & Bernhard 2021). In the context of L2 writing, female L2 learners, irrespective of L1 backgrounds, are proven to be overshadowing their male peers (Waskita 2008; Nia & Shahsavar 2019; Al-Saadi 2020). Therefore, a possible explanation for the low dimension score and the underuse of positive loading features of female learners might lie in their relatively higher language proficiency.

For PHL, one of the two exceptions that show an opposite trend against the other learner corpora (JPN is not discussed as the difference between its two genders on involvement is negligible), this study tends to assume that low involvement in its male learner corpus may have

Xiaoyun Li:

Involvement in Learner English writing: The Case from Asian Learners

Argumentum 18 (2022), 124–143

Debreceni Egyetemi Kiadó

DOI: 10.34103/ARGUMENTUM/2022/7

some aspects to link with the influence of native English (i.e., AmE). As shown in Appendix II, the male learner corpus of PHL contains less positive loading features but more negative loading features as compared to the female learner corpus, which echoes the following findings regarding the differences between the two genders of NES (mainly by American L1 writers) on the use of dimension 1 features.

The native corpus, NES, however, is in stark contrast to L2 learner corpora. The male corpus of it scores much lower than the female corpus does, which resonates with the research finding (Holmes 1995; Lakoff 1973; Biber & Burges 2000; Argamon et al. 2003) that female language production reveals a heavier use of features marking personal and interactional than does male language production. The result of the *t*-test also indicates that there is a statistically significant difference between the two genders of NES on involvement. Appendix II further confirms that the two genders of NES exhibit distinct preferences for linguistic features of dimension 1: female native writers, in comparison to their male counterparts, make more use of all positive features while the latter surpass them on all negative features. More specifically, female native writings can be described as interactive, involved, and interpersonal, whereas male native writings appear to be informational.

A question arises then: toward the same prompts, why are L1 writings significantly gendered regarding involvement while L2 writings are not? As a matter of fact, to say that gender differences are more significant in L1 writing than in L2 writing is not without research grounds among the rare literature focusing on gender-linked features in L2 writing. Samar & Shirazizadeh (2010) examined the gender-preferential features identified by Argamon et al. (2003) in L2 academic writing and found that the difference between the two genders in L2 production was not as evident as what was found by Argamon et al. (2003) in L1 production. They then proposed that the confinements from genres and non-nativeness are the two factors that keep “L2 writers from expressing their gender to its fullest capacity in the texts they produce” (Samar & Shirazizadeh 2010: 70). This study inclines to accept the latter factor since the language data utilized in the present study are highly homogenous regarding genre and are produced by L1 and L2 writers under closely controlled contexts. It can be expected that L2 writers may lack enough room to manifest their gender in writing as L1 writers do because L2 writing is recognized as “more constrained, more difficult, and less effective” (Silva 1993: 668) than L1 writing. Another possible explanation that might deserve noting is female learners' advantage over their male peers in language proficiency. More specifically, female language production may be more involved than male production as much earlier research indicated, but female L2 learners, owing to their advances in proficiency, consciously or unconsciously reduce the use of (or employ alternative non-involvement features to replace) features that might be typical in both female language and spoken language, thus narrowing the differences between two genders on involvement.

7 Conclusion

Under the guidance of three hypotheses drawing on the literature on involvement in English writings, this study probes into how Asian learners' non-native status, L1, language proficiency, and gender affect involvement in their written English production.

The results reveal that writings by Asian learners of English do not necessarily show higher involvement than those by native English speakers. It is found that learners from EFL contexts

generally do exhibit an overtly involved tendency, whereas learners from ESL contexts, in contrast, show a less involved style in writing than native English speakers. With regard to learners' L1, the results indicate that learners' L1 is not likely to affect the overall involvement in their writings, but it might exert influence on learners' usage on some specific involvement features. For the relationship between involvement and L2 learners' proficiency, the present study finds a negative correlation between involvement degree and learner proficiency, viz, involvement decreases as learners' English proficiency advances. Lastly, the results regarding the relationship between gender and involvement demonstrate that male and female native writers, as is concluded in earlier research, respectively reflect an informational and an involved focus and the difference is statistically significant. Nevertheless, for the Asian L2 learners of English, female learners exhibit lower involvement than their male peers though the difference is in general not statistically significant.

The reached results may carry important pedagogical implications. The low involvement discovered in learner writings from Asian ESL contexts, along with the relatively high involvement in writings by L1 writers may help overcome people's stereotype that learner language is characteristic of high involvement. One implication of this is that high involvement is better to be treated as the feature of novice writing due to writers' lack of register awareness because it can occur both in novice L1 and L2 writings and in dealing with involvement in writings, L1 writers are shown not in a superior position though they possess the perfect competence of English. As a result, raising register awareness is a necessity for writers of any language backgrounds who attempt to increase their writing ability.

To relate high involvement in writings to the lack of register awareness, however, does not mean that involvement features should be avoided or even eliminated in writing. This leads to another major pedagogical implication carried by the results of this study – the necessity of rejecting bias against involvement features. Compared with writings by experienced L1 writers, writings by advanced L2 learners in this study show an overly low involved tendency manifested mainly by avoiding typical involvement features like *first-* and *second-person pronouns* and *private verbs*. This might be problematic since involvement features, as mentioned before, perform important rhetoric functions. In discussing scholarly writing, Hyland (2002b) criticized the style guides and textbooks which advocate impersonal and faceless discourse and pointed out that it might do “a disservice” (Hyland 2002b: 351) to novice writers to frame them with these writing rules or conventions. Therefore, to prevent L2 learners from being overwhelmed by high involvement or high detachment (informational tendency), L2 writing instructors, in addition to raising their register awareness, also should impart knowledge about involvement features to them at the same time.

References

- Ackerley, K. (2008): Using comparable expert-writer and learner corpora for developing report-writing skills. In: Taylor, T. C., Ackerley, K. & Castello E. (eds.): *Corpora for University Language Teachers*. Bern: Peter Lang, 259–273.
- Aijmer, K. (2001): *I think* as a marker of discourse style in argumentative Swedish student writing. In: Aijmer, K. (ed.): *A wealth of English: Studies in honour of Göran Kjellmer*. Göteborg: Acta Universitatis Gothoburgensis, 247–257.

Xiaoyun Li:

*Involvement in Learner English writing: The Case from Asian Learners**Argumentum 18* (2022), 124–143

Debreceni Egyetemi Kiadó

DOI: 10.34103/ARGUMENTUM/2022/7

- Ädel, A. (2008): Involvement features in writing: Do time and interaction trump register awareness? In: Gaëtanelle, G., Papp, S. & Diez-Bedmar, M.B. (eds.): *Linking up Contrastive and Learner Corpus Research Amsterdam*. NY: Rodopi, 35–53.
https://doi.org/10.1163/9789401206204_003
- Altenberg, B. (1997): Exploring the Swedish component of the International Corpus of Learner English. In: Lewandowska-Tomaszczyk, B. & Melia, P. J. (eds.): *Practical Applications in Language Corpora*. Łódź: Łódź University Press, 119–132.
- Al-Saadi, Z. (2020): Gender differences in writing: The mediating effect of language proficiency and writing fluency in text quality. *Cogent Education* 7.1, 1–19.
<https://doi.org/10.1080/2331186X.2020.1770923>
- Alhusban, H. A. & Vijayakumar, C. (2021): Lexical Bundles in Saudi EFL Student Writing: A Study of Learner Corpus. *TESOL International Journal* 16.4.2, 7–31.
- 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>
- Barbieri, F. (2013): Involvement in university classroom discourse: register variation and interactivity. *Applied Linguistics* 36.2, 151–173. <https://doi.org/10.1093/applin/amt030>
- Bednarek, M. (2014): Involvement in Australian Talkback Radio—A Corpus Linguistic Investigation. *Australian Journal of Linguistics* 34.1, 4–23.
<https://doi.org/10.1080/07268602.2014.875453>
- Bernhard, S. & Bernhard, S. (2021): Gender differences in second language proficiency—Evidence from recent humanitarian migrants in Germany. *Journal of Refugee Studies* 35.1, 282–309. <https://doi.org/10.1093/jrs/feab038>
- Biber, D. (1987): A textual comparison of British and American writing. *American speech* 62.2, 99–119. <https://doi.org/10.2307/455273>
- Biber, D. (1988): *Variation across speech and writing*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511621024>
- Biber, D. & Burges, J. (2000): Historical change in the language use of women and men: Gender differences in dramatic dialogue. *Journal of English linguistics* 28.1, 21–37.
<https://doi.org/10.1177/00754240022004857>
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (eds.) (1999): *Longman grammar of spoken and written English*. London: Longman. <https://doi.org/10.1075/z.232>
- Chafe, W. & Danielewicz, J. (1987): Properties of spoken and written language. In: Horowitz, R. & Samuels, S. J. (eds.): *Comprehending oral and written language*. New York: Academic Press, 83–113.
- Chafe, W. (1985): Linguistic differences produced by differences between speaking and writing. In: Olson, D.R., Torrance, N. & Hildyard, A. (eds.): *Literacy, Language, and Learning: the nature and consequences of reading and writing*. Cambridge: Cambridge University Press, 105–123.
- Cobb, T. (2003): Analyzing later interlanguage with learner corpora: Quebec replications of three European studies. *Canadian Modern Language Review* 59.3, 393–423.
<https://doi.org/10.3138/cmlr.59.3.393>
- Crosthwaite, P. (2016): A longitudinal multidimensional analysis of EAP writing: Determining EAP course effectiveness. *Journal of English for Academic Purposes* 22, 166–178.
<https://doi.org/10.1016/j.jeap.2016.04.005>
- Evans, S. (2011): Hong Kong English and the professional world. *World Englishes* 30.3, 293–316. <https://doi.org/10.1111/j.1467-971X.2011.01655.x>

Xiaoyun Li:

*Involvement in Learner English writing: The Case from Asian Learners**Argumentum 18 (2022), 124–143**Debreceni Egyetemi Kiadó*

DOI: 10.34103/ARGUMENTUM/2022/7

- Friginal, E., Li, M. & Weigle, S. C. (2014): Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing* 23, 1–16. <https://doi.org/10.1016/j.jslw.2013.10.001>
- Gilquin, G. & Paquot, M. (2007): Spoken features in learner academic writing: Identification, explanation and solution. In: *Proceedings of the 4th Corpus Linguistics Conference*, University of Birmingham, 27–30.
- Gilquin, G. & Paquot, M. (2008): Too chatty: Learner academic writing and register variation. *English Text Construction*, 1.1, 41–61. <https://doi.org/10.1075/etc.1.1.05gil>
- Główka, D. (2014): The impact of gender on attainment in learning English as a foreign language. *Studies in Second Language Learning and Teaching* 4.4, 617–635. <https://doi.org/10.14746/ssl1t.2014.4.4.3>
- Grant, L. & Ginther, A. (2000): Using computer-tagged linguistic features to describe L2 writing differences. *Journal of second language writing* 9.2, 123–145. [https://doi.org/10.1016/s1060-3743\(00\)00019-9](https://doi.org/10.1016/s1060-3743(00)00019-9)
- Harwood, N. (2005): ‘Nowhere has anyone attempted... In this article I aim to do just that’: A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics* 37.8, 1207–1231. <https://doi.org/10.1016/j.pragma.2005.01.012>
- Holmes, J. (1995): *Women, men, and politeness*. London: Longman. <https://doi.org/10.4324/9781315845722>
- Hu, Y. & Bodomó, A. B. (2009): Harbilingual: L1 influence on the learning of English by high school students in Harbin, China. *The Asian EFL Journal Quarterly* 11.3, 190–233.
- Hubbard, E. H. (2001): Interaction as ‘involvement’ in writing for students: a corpus linguistic analysis of a key readability feature. *Southern African linguistics and applied language studies* 19, 231–240. <https://doi.org/10.2989/16073610109486289>
- Hyland, K. (2002a): Authority and invisibility: Authorial identity in academic writing. *Journal of pragmatics* 34.8, 1091–1112. [https://doi.org/10.1016/S0378-2166\(02\)00035-8](https://doi.org/10.1016/S0378-2166(02)00035-8)
- Hyland, K. (2002b): Options of identity in academic writing. *ELT journal* 56.4, 351–358. <https://doi.org/10.1093/elt/56.4.351>
- Lingley, D. (2005): Spoken features of dialogue journal writing. *Asian EFL Journal* 7.2, 48–67.
- Issit, S. (2016): *Evaluating the impact of a professional English for academic purposes program: a corpus-based study*. Doctoral dissertation from University of Birmingham. <https://theses.bham.ac.uk/id/eprint/7080/1/Issitt17PhD.pdf> (Accessed 15 January 2021)
- Ishikawa, S. (2013): The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. *Learner corpus studies in Asia and the world* 1, 91–118.
- Ishikawa, S. (2014): Design of the ICNALE-Spoken: A new database for multi-modal contrastive interlanguage analysis. *Learner corpus studies in Asia and the world* 2, 63–76.
- Kobayashi, Y. & Abe, M. (2016): A corpus-based approach to the register awareness of Asian learners of English. *Journal of Pan-Pacific Association of Applied Linguistics* 20.2, 1–17.
- Lakoff, R. (1973): Language and woman’s place. *Language in Society* 2.1, 45–79. <https://doi.org/10.1017/S0047404500000051>
- Lee, J. J., Bychkovska, T. & Maxwell, J. D. (2019): Breaking the rules? A corpus-based comparison of informal features in L1 and L2 undergraduate student writing. *System* 80, 143–153. <https://doi.org/10.1016/j.system.2018.11.010>
- Martin, P. I. (2014): Philippine English revisited. *World Englishes* 33.1, 50–59. <https://doi.org/10.1111/weng.12054>

- Natsukari, S. (2012): Use of *I* in essays by Japanese EFL learners. *JALT Journal* 34.1, 61–78. <https://doi.org/10.37546/JALTJJ34.1-4>
- Nia, S. K. & Shahsavar, Z. (2019): Comparing the use of different prewriting strategies on medical students' L2 Writing. *The Asian ESP Journal* 15.3, 228–255.
- Nini, A. (2019): The Multi-Dimensional Analysis Tagger. In: Berber, S. T. & Veirano, P. M. (eds.): *Multi-Dimensional Analysis: Research Methods and Current Issues*. London; New York: Bloomsbury Academic, 67–94. <https://doi.org/10.5040/9781350023857.0012>
- Paquot, M., Hasselgård, H. & Ebeling, S. O. (2013): Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. In: Granger, S., Gilquin, G. & Meunier, F. (eds.): *Twenty years of learner corpus research: Looking back, moving ahead*. Louvain-la Neuve: Presses universitaires de Louvain, 377–387.
- Petch-Tyson, S. (1998): Writer/reader visibility in EFL written discourse. In: Granger, S. (eds.): *Learner English on Computer*. London & New York: Longman, 107–118. <https://doi.org/10.4324/9781315841342-8>
- Read, J. (2000): *Assessing Vocabulary*, Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
- Recski, L. J. (2004): Expressing standpoints in EFL written discourse. *Revista Virtual de Estudos da Linguagem* 2.3, 1–16.
- Samar, R., & Shirazizadeh, M. (2010): Gender-preferential linguistic elements in applied linguistics research papers: Partial evaluation of a model of gendered language. *The Journal of Applied Linguistics* 3.1, 70–84
- Silva, T. (1993): Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly* 27.4, 657–677. <https://doi.org/10.2307/3587400>
- Tannen, D. (1983): Oral and literate strategies in spoken and written discourse. In: Bailey, R.W. & Fosheim, R.M. (eds.): *Literacy for life: The demand for reading and writing*. New York: Modern Language Association of America, 79–96.
- Waskita, D. (2008): Differences in men's and women's ESL academic writing at the University of Melbourne. *Jurnal Sosioteknologi* 7.14, 448–463.
- Wolfe-Quintero, K., Inagaki, S. & Kim, H.Y. (eds.) (1998): *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.
- Yoon, C. (2015): Writing like speaking: spoken register features in Korean EFL learners' argumentative writing. *Korean Journal of Applied Linguistics* 31.4, 153–179. <https://doi.org/10.17154/kjal.2015.12.31.4.153>

Xiaoyun Li
 University of Szeged
 Faculty of Humanities and Social Sciences,
 Doctoral School in Linguistics, Theoretical Linguistic Programme
 H-6721 Szeged, Tisza Lajos krt. 36.
 lixiao@hung.u-szeged.hu

Appendix I: The frequencies of Dimension 1 features of the 11 corpora
 (Per 1000 words)

	ESL Learner Groups					EFL Learner Groups					
	ENS	PHL	PAK	SIN	HKG	KOR	JPN	TWN	THA	IDN	CHN
Features with positive loadings											
AMP	4.6	3.3	7.1	4.7	4.9	6.4	7	5.6	5	4.3	5.3
ANDC	6.1	3.7	2.6	3.9	2.7	5	6.3	5.7	4	5.6	4.4
BEMA	24.7	24.2	24.2	23.9	25.3	29.2	23.6	24	22.2	20.4	24.9
CAUS	3.6	6.3	3.7	2.2	2.2	5.4	4.2	3.6	8.1	9	2
CONT	6.2	7.3	2.1	0.8	2.2	6.8	8.6	8.9	7.8	4.9	8
DEMP	7.9	5.9	3.8	6.7	4.5	4.5	4.7	3.9	3.8	4.7	3.3
DPAR	0.2	0.3	0.9	0	0	0.3	0.5	0.3	0.4	0.4	0.4
EMPH	12.9	10.7	6.6	7.9	10.8	10.8	9.1	14.2	9.8	9	16
FPP1	33.4	23.5	12.6	14.8	16.1	30.4	50.8	35.6	24.4	25.8	38
HDG	0.4	0.5	0	0.1	0.1	0.3	0.2	1.2	0.4	0.7	0.7
INPR	0.5	0.1	0.3	0.1	0.4	0.3	0.3	0.2	0.2	0.2	0.3
PIT	15.2	19.6	18.9	14.8	14	13.7	14.1	14.5	17.2	13.5	16.6
POMD	10	14.8	10.4	12.8	16.6	13.6	13.4	17.1	15.7	18.4	17.6
PRED	12.8	12.2	12.5	15	14.3	16.8	15.6	13.4	11.7	10.4	13.9
PRIV	21.5	13.1	7.5	14	12.9	19	22.2	22.5	16.8	13.1	19
PROD	2.6	2.1	7.3	1.1	2.1	3.3	2.8	3	5	3.5	4.2
SERE	0.3	0.4	0.3	1.1	1	0.3	0.1	0.6	0.3	0.3	1.4
SPP2	5.7	12.9	2	1.5	4.4	12	5.8	12.2	19.2	6	9.9
STPR	0.9	0.7	0.2	0.8	0.8	0.8	1.3	0.9	0.6	0.3	1.1
THATD	3.5	1.8	0.7	1.6	2.4	4.2	4.8	5.2	3.5	2.1	3.8
VPRT	75	74.4	83.2	61.1	67.7	86.8	95.5	78.1	78.5	72.3	76.1
WHCL	0.9	1.2	0.2	0.6	0.8	1.1	1.5	2.2	0.9	1.1	1.1
WHQU	0.3	0.3	0.1	0.1	0.5	0.3	0.2	0.5	0.3	0.1	1.1
XX0	13.3	14.4	12.6	10.8	12.8	15.3	17.6	17.3	16.7	16.5	13.5
Features with negative loadings											
AWL	4.46	4.51	4.48	4.79	4.76	4.53	4.38	4.43	4.38	4.5	4.46
TTR	126.5	125.7	116.6	130.3	122.7	113.2	108.2	118.4	109.5	111.2	124.1
JJ	60.4	49.9	60.1	64.4	65.3	57.2	47.8	56.6	55.3	53.7	60.8
NN	205.5	204.2	243.9	201.8	211.6	221.5	201	198.4	213.3	230.9	197.4
PIN	90.4	91.8	98.9	100.5	94.7	83.5	88.3	80.2	83.1	88.3	91

Appendix II: The frequencies of Dimension 1 features of the two genders
 (Per 1000 words)

	NES		PHL		PAK		SIN		HKG		KOR	
	F	M	F	M	F	M	F	M	F	M	F	M
Positive features												
AMP	5.2	4.2	3.0	3.7	7.5	5.8	4.8	4.7	5.9	4.4	6.6	6.1
ANDC	6.2	6.0	3.6	3.7	2.5	2.7	3.6	4.5	2.6	2.8	4.2	5.1
BEMA	26.2	23.8	24.6	23.6	24.4	23.6	24.8	22.5	26.5	24.8	28.7	29.1
CAUS	3.9	3.4	5.7	7.1	4.0	2.8	2.3	2.0	1.9	2.3	4.9	5.9
CONT	6.6	6.0	5.1	7.1	1.8	3.3	0.4	1.4	2.4	2.2	7.1	7.1
DEMP	8.6	7.5	6.2	5.5	3.5	4.6	6.5	7.0	3.9	4.8	4.1	5.3
DPAR	0.2	0.2	0.3	0.3	1.0	0.7	0	0.1	0	0	0.3	0.3
EMPH	13.9	12.3	11.2	10.0	6.0	8.5	7.9	8.0	11.4	10.5	10.2	11.2
FPP1	36.4	31.6	23.7	23.2	12.1	14.2	15.6	13.5	17.6	15.5	30.0	30.8
HDG	0.5	0.4	0.5	0.3	0	0	0	0.2	0.2	0.1	0.4	0.2
INPR	0.5	0.5	0.1	0.2	0.3	0.5	0.1	0.2	0.4	0.4	0.3	0.5
PIT	16.4	14.5	19.2	20.2	18.1	21.5	14.7	14.9	13.5	14.2	14.2	13.6
POMD	10.8	9.4	15.6	13.6	10.8	9.3	12.6	13.2	18.3	15.8	14.7	12.1
PRIV	23.6	20.1	13.3	12.9	7.2	8.4	14.2	13.7	12.8	13.0	19.2	18.6
PROD	2.7	2.5	2.2	2.0	7.0	8.1	0.9	1.3	1.8	2.3	3.1	3.3
SERE	0.3	0.4	0.3	0.5	0.3	0.3	1.1	1.1	0.5	1.2	0.3	0.4
SPP2	6.8	5.0	13.3	12.3	1.3	4.1	1.0	2.3	3.5	4.7	9.4	16.5
STPR	0.7	1.1	0.7	0.8	0.2	0.2	0.8	1.0	0.8	0.9	0.7	1.1
THATD	3.9	3.3	1.9	1.7	0.7	1.0	1.5	1.8	2.0	2.6	3.7	4.8
VPRT	76.7	74	74.2	74.6	82.6	84.9	60.9	61.6	66.8	68.0	87.8	85.9
WHCL	1.2	0.7	1.2	1.1	0.3	0.1	0.6	0.5	1.2	0.7	1.4	0.9
WHQU	0.3	0.3	0.2	0.3	0.1	0.1	0.1	0.1	0.6	0.4	0.3	0.4
XX0	14.3	12.7	14.6	14.3	13.2	10.8	10.8	11.0	12.1	13.1	14.9	16.0
Features with negative loadings												
AWL	4.37	4.52	4.49	4.54	4.5	4.39	4.79	4.77	4.82	4.73	4.55	4.49
TTR	122.4	129	120.1	121.5	115.1	121.1	129.5	131.6	123.8	122.3	111.8	114.7
JJ	57.7	61.9	48.6	51.9	59.6	61.7	64.4	64.3	65.8	65.1	57.2	58.0
NN	166.4	181.2	203.3	205.7	242.6	247.9	201.8	201.9	208.4	212.9	225.3	218.4
PIN	88.2	91.9	91.1	92.9	99.6	96.8	101.9	98.0	92.9	95.4	82.2	82.9

Appendix II: The frequencies of Dimension 1 features of the two genders (Cont.)

	JPN		TWN		THA		IDN		CHN	
	F	M	F	M	F	M	F	M	F	M
Features with positive loadings										
AMP	6.5	7.3	5.4	6.2	5.1	4.8	4.1	4.5	5.6	4.8
ANDC	5.9	6.5	5.5	6.0	4.0	4.0	5.5	5.8	4.1	4.7
BEMA	23.2	23.8	23.6	25.0	22.8	21.2	20.0	20.7	24.4	25.3
CAUS	4.1	4.2	3.5	4.0	8.3	7.7	7.9	10.0	2.0	1.9
CONT	9.6	8.1	9.0	8.5	6.9	9.2	4.5	5.3	7.5	8.5
DEMP	4.4	4.9	3.9	3.9	3.7	4.2	5.2	4.3	3.3	3.5
DPAR	0.4	0.5	0.2	0.5	0.4	0.4	0.2	0.6	0.4	0.4
EMPH	9.2	9.0	14.5	13.4	9.1	11.0	9.1	8.8	15.5	16.6
FPP1	50.8	50.8	36.3	33.6	23.0	27.0	25.9	25.8	37.8	37.7
HDG	0.2	0.2	1.1	1.3	0.4	0.4	0.7	0.8	0.5	0.9
INPR	0.4	0.3	0.2	0.3	0.3	0.2	0.3	0.1	0.2	0.3
PIT	14.1	14.1	14.6	14.0	17.1	17.6	13.0	13.9	16.5	16.5
POMD	15.1	12.4	16.9	17.8	15.8	15.5	18.1	18.6	17.5	17.5
PRIV	23.4	21.6	22.0	23.9	16.3	17.5	12.3	13.9	18.6	19.2
PROD	2.5	3.0	2.9	3.4	4.7	5.7	3.5	3.5	3.9	4.6
SERE	0.1	0.1	0.6	0.4	0.3	0.2	0.2	0.3	1.4	1.5
SPP2	5.6	5.9	10.4	17.6	20.8	16.4	4.9	7.0	9.7	9.9
STPR	1.2	1.3	1.0	0.6	0.6	0.5	0.3	0.4	1.2	1.0
THATD	5.0	4.6	5.2	5.1	3.3	3.8	2.1	2.1	3.5	4.1
VPRT	92.4	97.2	77.8	79.0	78.4	78.6	71.1	73.4	76.0	76.2
WHCL	1.7	1.4	2.2	2.2	0.8	1.0	1.1	1.2	1.0	1.2
WHQU	0.2	0.2	0.6	0.4	0.3	0.2	0.1	0.2	1.1	1.2
XX0	17.4	17.7	18.1	15.1	16.5	17.1	15.5	17.5	12.6	14.2
Features with negative loadings										
AWL	4.39	4.37	4.43	4.45	4.4	4.34	4.52	4.48	4.5	4.42
TTR	115.4	106.4	118.1	119.2	109.4	109.7	115.2	107.6	124.7	124.1
JJ	47.7	47.8	56.3	57.4	56.2	53.8	54.6	53.0	61.6	60.4
NN	197	203.2	198.5	198.1	216.7	207.4	231.0	230.8	198.2	196.7
PIN	90.2	87.2	80.2	80.1	82.9	83.3	87.9	88.8	92.6	89.9