A comparative study of abstraction and informational density in higher and lower Indonesian graduate students’ L2 English academic texts

DEDY SUBANDOWO

Department of Theoretical Linguistics, Pázmány Péter Catholic University, Budapest, Hungary

RESEARCH ARTICLE

Received: March 1, 2022 • Accepted: May 17, 2022
Published online: June 27, 2022
© 2022 The Author(s)

ABSTRACT

The language features of the second language (L2) English academic texts written by Indonesian graduate students enrolled in Hungarian higher education are employed in the present study. The study focuses on the level of abstraction and informational density in student assignments in particular. Seven high-stakes essays were collected from seven Indonesian graduate students registered in the faculty of Social Sciences at three different Hungarian universities. Coh-Metrix, a corpus-based computational tool, was used in this study to examine indices of content words and abstraction. As for the comparison between higher and lower proficiency level students, parametric statistics were used to conduct a quantitative analysis of the selected linguistic features, including nouns, verbs, adjectives, adverbs, pronouns, and concreteness and abstraction. The results show that C1 English proficiency level students outnumber B2 students in terms of informational density. Their texts are more abstract than those of B2 students.

KEYWORDS

abstraction, academic writing, Hungarian higher education, lexical density, second language writing, writing improvement

* Corresponding author. E-mail: subandowo.dedy@hallgato.ppke.hu
INTRODUCTION

It has been the fact that internationalization sounds very positive about global education, particularly higher education. This primary objective is relevant to the purpose of the Stipendium Hungaricum scholarship program, which was established in 2013 by the Hungarian Government in order to promote Hungarian higher education around the world as well as to attract top international students from all continents who could establish personal and professional attachments to Hungary while receiving a high-quality education in the heart of Europe (Web1, https://stipendiumhungaricum.hu/about/, 2020). This program, based on bilateral education agreements between Hungary and the governments of the sending countries, is already available in nearly 80 countries across five continents, including Indonesia.

In the early stages of the agreement between the two countries in 2016, the number of scholarships available was limited to 50 students. Meanwhile, Indonesia has seen its GDP more than double since 2020. According to data provided by the Indonesian Students Association in Hungary, the total number of students enrolled at Hungarian universities exceeds 200 (Gariahub, 2021). As a result of this circumstance, English has become essential for students to communicate theoretically. As a result, English as a medium of instruction is an inherent part of education, particularly when students study in a non-native English-speaking environment (Macaro, Curle, Pun, An, & Dearden, 2018).

As international students in Hungarian higher education, they must meet the university’s English proficiency requirements. In most cases, Hungarian universities require a B2 or higher (or 6.0 IELTS) for a master’s program in the Social Science faculty. Nonetheless, the English proficiency certificate requirement is not the only high-stakes test for students’ success. Singh (2015) discovered that the stakes are high in the taught Master’s programs in which international graduate students enroll. English as a medium of instruction may pose academic challenges for non-native English students. Hyland (2006) also remarked that most graduate students would be regularly exposed to lectures, seminars, and exams and be required to take notes, give presentations, and writing assignments. Furthermore, the latter issue considers being relevant to academic writing.

Students may face enormous challenges in language functions, including linguistic features, in academic writing. Linguistic features are essential in indicating text difficulty and quality in academic writing. Texts with higher scores were more likely to have linguistic features associated with sophisticated language (McNamara, Crossley, & McCarthy, 2010). Furthermore, Swales (1995) and Staples, Egbert, Biber, and Gray (2016) identified that informational density and nominalization appear to be inherent linguistic features.

THEORETICAL BACKGROUND

English as a medium of instruction in higher education (EMI)

The use of English as a lingua franca (ELF) has been the main means of communication in academic settings worldwide due to the acceleration of globalization (Crystal, 2003). These circumstances have significant relations with the increased introduction of English as a medium of instruction (EMI) in tertiary education in many countries worldwide. Dearden (2016) defines “EMI as the use of the English language to teach academic subjects in countries or Jurisdictions
where the first language (L1) of the majority of the population is not English” (p. 4). The definition points out that EMI in an English L1 environment or English-dominant environment may not be considered EMI due to the use of English as an L2.

EMI is being used by many tertiary institutions to an increasing extent, including universities (see, e.g., H. Coleman, 2011; J. Coleman, Hultgren, Li, Tsui, & Shaw, 2018; Macaro et al., 2018). It aims to promote students’ and faculty members’ mobility to make them competitive and employable in international settings and achieve attractive and reputable institutions. In addition, Macaro et al. (2018) explain that the growing phenomena of EMI in higher education establish obvious connections with institutional aspects involving a perceived internationalization, foreign student attraction due to the decreasing enrolment number of home students through demographical changes, national cuts in higher education investment, the competition between state and private sectors, and the use of English as the international language, expressly in the domain of research publications.

It has become the fact that more and more universities are willing to offer both undergraduate and graduate programs in English (Doiz, Lasagabaster, & Sierra, 2014). One study by Björkman (2011) examined the increasing use of English as a lingua franca as the medium of instruction for higher education in continental Europe and elsewhere. The study reported on the new group of learners who predominantly need the medium of English to communicate with speakers from other first language backgrounds. Another study by Petzold and Berns (2000) investigated that English is increasingly used in Hungary as an effective medium in education and has a major impact on learning, especially in Budapest higher education. The study found that students at the university have had contact with English through up-to-date information by reading various texts, including professional journal articles, reference works, and textbooks.

**English academic writing**

Writing is a key skill for students to learn and express what they confirm. Writing is an essential skill that all language learners should cultivate (Baghbadorani & Roohani, 2014). Writing is a key aspect at universities, and it also presents significant challenges to second language (L2) learners. According to Shoﬁya (2013), writing is the most difficult language skill to learn when learning a second language. Writing is usually required in school for papers, reports, and theses, and it plays an important role in teaching and learning (Chien, 2012, p. 93). Writing is also a thinking tool that allows students to express themselves and understand and share their perspectives on the world around them (Clark, 2014, p. 6).

Writing development is necessary for university students because it has become a critical skill for their academic development. Writing is an important skill for knowledge production and dissemination in any disciplinary discourse (Raoofi, Binandeh, & Rahmani, 2017). Assessing students’ academic achievement in academic contexts is heavily reliant on their ability to communicate their language knowledge and ideas. It assists students in completing key assignments, improving critical thinking skills, and developing cognitive performance and function (Graham & Perin, 2007). In an educational context, the ability to write in English is critical for university students to function successfully in their academic areas. Writing in English in academic circles allows students to share their research findings with global readerships and place their thoughts and research in an international outlet (Raoofi et al., 2017).

Canagarajah (2002) emphasizes the significance of organizing writing into five key features. First, writing reflects and creates reality; second, writing is a social interactional activity between
the writer and the reader within a specific space and time. Third, writing results from a negotiation between the writer and the available resources in a given context. Fourth, writing allows writers to present ideological beliefs, express themselves, and assign value to entities through the text. Fifth, writing is a historical dynamics process in which the ideas, struggles, conflicts, and concepts of the text are open to the comments and stances of the readers and writers.

Writing is now regarded as more than just a means of communication in academic circles. As a result, conveying meaning in written texts is critical for academic and professional success (Dastjerdi & Samian, 2011). Geiser and Studley (2002) stated that the writing abilities of first-year college students are among the best predictors of academic success. Many L2 learners find the writing process difficult because they must make sense of and reflect on their thoughts on specific topics (Zimmerman & Risemberg, 1997). To achieve academic success, students must employ writing strategies.

Writing has been described as having three main activities: planning, formulating or composing, and revising, which in the traditional understanding of writing was understood as a linear procedure, a strict 'plan-outline-write' that had little to do with the complex activities that teachers observed in the composting process of their writers, as these were much more than building grammatically correct sentences (De La Paz, 2009; So & Lee, 2013; Winarto, 2015). As a direct consequence, writing demonstrates students' ability to master writing techniques, so students must be aware of writing as both a process and a product (Okasha & Hamdi, 2014). Students should not only be able to control and manage their writing skills, rules, and conventions, but they should also be able to apply some strategies to successfully finalize their writing. According to research, effective and successful writers use a variety of writing strategies to construct their paragraphs (e.g., Alias, Manan, Yusof, & Pandian, 2012; Torrance & Thomas, 1994). It appears to mean that writers have goals and tailor their writing to each goal and writing task.

L2 English academic writers

The term “second language writer” refers to anyone who is writing or learning to write in a language other than their native language (Matsuda & Silva, 2020). It consists of both second and foreign language writers and writers who are writing in their third, fourth, and fifth language, and so forth. Second language writers seem to encounter some constraints, including language, strategy, and culture (Hyland, 2003). They do not only learn to write, but second language writers acquire second language structures (Matsuda & Jablonski, 2000). Apart from the second language grammar, L2 writers also need to develop their language proficiency and genre knowledge.

Another definition of second language writers classified by Matsuda and Jablonski (2000) refers to foreign language writers. It indicates writers who are writing in languages in which the target language is not popular. This type of writer can be grouped into two general categories: EFL writers and other foreign language writers. It becomes evident that the distinction between EFL and other foreign languages is significant even if they investigate the same academic unit. The main reason for the distinction is based on the status of the English language as the dominant language of universal communication, especially in academic discourses. As a result, with the dominance of English as a lingua Franca of scholarly communication, writing in academic contexts for graduate students and researchers has been a major emphasis in many applied linguistics studies.
Linguistic features in L2 English academic writing

Writing is an essential constituent of higher education. It greatly helps students in communicating their ideas and promotes academic success. On the other hand, writing an academic text appears to be challenging for English learners. They must learn to write in a second language in addition to mastering grammatical structures and amassing sizable vocabularies, as well as make a distinction between academic and conversational English in terms of conventional words, phrases, and sentence structures (register knowledge; Biber & Conrad, 2019). Students must therefore write in their academic disciplines using specific linguistic features.

Writing an academic text is strongly associated with the educational genre in the context of higher education. According to Hyland (2003b, p. 18), a genre is a way of using language for specific purposes. He divides academic genres into subtopics, one of which is student essays (see Hyland, 2008, p. 11). Besides that, Crossley (2020) argued that students’ writing assignments could be used to assess linguistic features.

Fang (2005) investigated some characteristics of scientific writing regarding particular linguistic features. The findings show that lexical density and abstraction are classified as special features. The proportion of lexical richness in a text is its lexical density (Gregori-Signes & Clavel-Arroita, 2015). This feature can then be used to gauge the progress of students’ writing. On the other hand, abstraction refers to language features presented by the writer to abstract assumptions from everyday perceptions of the world (Hyland, 2009).

Informational density and abstraction in academic texts

It is undeniably true that non-native English speakers appear to face greater difficulties in writing English texts. Writing a high level of English is a vital issue concerning the challenge. Linguistic features, for example, make a substantial contribution to the quality of writing. Examining writing quality is critical to determining vocabulary richness. Text richness is related to lexical density because it focuses on the proportion of content words in a text (Laufer & Nation, 1995). Laufer and Nation (1995) define lexical density as the proportion of content words in a text: nouns, verbs, adjectives, or adverbs. According to Halliday, Lexical density is the type of complexity typical of written language (1989:62).

Furthermore, according to Read (2007), lexical density represents a proportion of lexical items in the form of information and ideas in written texts. Furthermore, abstraction is a linguistic feature that contributes to academic writing (Fang, 2004). It is expected that informational density and abstraction could then be used to assess students’ writing progress.

There are many research studies evaluating English academic texts (see, e.g., Crossley, 2020; Fang, 2005; MacIntyre, 2019; Russell, 2014). The majority of studies aimed to examine linguistic features as an indicator of academic text quality. According to Biber (1996), linguistic features in academic writing are more specific and technical than in spoken interactions. The use of content words in academic texts demonstrates the richness of lexical density in this case. In this sense, the density of a standard text, such as academic writing, is greater than that of an informal text (see, e.g., Biber, 1996). Similarly, research studies have shown that the lexical density in abstracts is high because they are condensed and have a word limit that allows writers to express key ideas clearly (see, e.g., the discussion in Nasseri & Thompson, 2021).

A number of studies focusing on lexical density have piqued the interest of applied linguists. The results show that higher English proficiency levels and English L1 texts have higher lexical
density values. For example, Nasseri and Thompson (2021) investigated differences in lexical density and abstract thesis in English L1 and L2 (both EFL and ESL) academic writing. The study found that EFL students produced the least lexically dense and diverse texts compared to ESL students, whereas L1 English students produced more sophisticated phrasal complexity, resulting in higher informational density in their texts. Kim (2006) said that lexical density is a strong predictor of academic writing proficiency. In a cross-sectional study, he also analyzed a corpus of college essays written by Korean students to distinguish proficiency levels based on the CEFR. Similarly, Gregori and Clavel (2015) discovered that texts written by students with higher levels of English proficiency have a high lexical density. Linnarud (1975) discovered that texts written by L1 English speakers have a higher level of lexical density than those written by L1 Swedish speakers.

Another area of focus in academic writing research is an abstraction. Language features in academic texts differ from the language used to describe everyday life experiences. Academic writing has a strong desire to be nominal (e.g., Biber, 1996; Biber & Gray, 2016). These characteristics distinguish academic writing from a conversation and fiction, making reading academic writing difficult for college students, whether graduate or undergraduate (Parkinson, 2020). Abstract nouns are frequently found in the writer registers, which account for 50% of all nouns, according to Biber (1996). Similarly, Syarif and Putri (2018) argue that a higher proportion of nouns in academic texts is important in increasing informational density.

According to Biber and Gray (2016), linguistic features in the humanities usually involve highly specialized vocabularies like abstraction. According to Hyland (2009a: 7), the writing process involves assumptions made by both the writer and the reader. This assumption in academic writing could be considered abstract. The abstraction makes it difficult for non-specialist readers to understand the texts’ contents. Furthermore, Fang (2005) pointed out that abstraction in a scientific text concerns nominalization. Nominalization is transforming verbs or adjectives into participants expressed by nouns. This abstraction is regarded as one of the unique features of scientific texts and a particularly powerful resource for synthesizing or abstracting information into entities for subsequent discussion.

Aims and research questions

The prime aim of this present study is to compare abstraction and informational density based on the proficiency levels of Indonesian graduate students enrolled in Hungarian higher education. In order to pursue the aim, two main questions are formulated as follows:

1. What is the relationship between students’ level of proficiency and the level of information density in their academic essays?
2. What is the relationship between students’ level of proficiency and abstraction in their academic essays?

METHODS

The present study was descriptive, comparative, and applied quantitative, statistical analysis with SPSS and Corpus-based analysis platforms.
Sample

This preliminary study’s analysis was based on seven essays written by graduate students pursuing Master’s degrees in the Social Science faculty at three major universities in Hungary. Data were collected via email after students submitted their essay assignments and received feedback and score provided by their professor at the end of each semester. The assignments were written at home with the deadline given by the professor. The assignments were based on their disciplines discussed with their lecturer. Most of the topic was relevant to the field of economic studies. The essay assignment was a high-stakes piece of writing submitted and graded by the professors. Table 1 shows the total number of words, sentences, and paragraph lengths submitted by the students during the two semesters of the study.

The students who participated in the study were 27–29 years old, and they were divided into two groups based on their former IELTS proficiency levels: three students with higher proficiency levels (C1; 7.0–7.5) and four students with lower proficiency levels (B2; 6.0–6.5) when enrolled to the university. The students were registered in the three core disciplines, including the Department of International Relations, Social Integrations, and Regional and Environmental Economics. The sample was selected from different universities due to a limited number of participants who enrolled in mentioned disciplines and provided equivalence distribution across the study in the Hungarian higher education.

The Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) was used in the analysis to assess the presence of linguistic features attributed to the frequency of informational density and abstraction. The informational density is the number of content words per non-embedded clause or the percentage of content words relative to total running words (Fang, 2004). On the other hand, the abstraction includes several imageability measures, which are specifically related to mental image construction in one’s mind (Graesser et al., 2004).

Figure 1 demonstrates Coh-metrix’s automated analysis with a descriptive explanation of the uploaded text. The text is analyzed based on narrativity, syntactic simplicity (e.g., simple sentence), word concreteness (e.g., abstraction, informational density), referential cohesion, and deep cohesion (e.g., discourse markers). These category analyses refer to text ease, which means the easier to read the text is, the more concrete the text is.

<table>
<thead>
<tr>
<th>Disciplines</th>
<th>English proficiency levels (IELTS)</th>
<th>Number of texts</th>
<th>Number of words</th>
<th>Number of sentences</th>
<th>Paragraph length</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Relations, Regional and Environmental Economics</td>
<td>C1</td>
<td>3</td>
<td>9,420</td>
<td>494</td>
<td>3,362</td>
</tr>
<tr>
<td>International Relations, Social Integration, and Regional and Environmental Economics</td>
<td>B2</td>
<td>4</td>
<td>10,469</td>
<td>678</td>
<td>2,986</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>7</td>
<td>19,889</td>
<td>1,173</td>
<td>6,348</td>
</tr>
</tbody>
</table>
As for data analysis, the first step was to sort submitted and marked texts based on the defined classification of students’ higher and lower English proficiency levels. The texts are uploaded to the Coh-Metrix 3.0 webool (Tackett & McNamara, n.d. http://www.cohmetrix.com/), and the results are downloaded directly.

RESULTS AND DISCUSSION

Seven essays written by Indonesian graduate students from three different disciplines were evaluated, and the results were used to determine the frequency of linguistic features defined by their English proficiency levels. The emphasis of linguistic features is on information density and abstraction. Informational density is assessed for nouns, verbs, adjectives, adverbs, and pronouns. In contrast, abstraction emphasizes concreteness and imagibility. Statistical analysis confirms that the data are normally distributed ($P$-value $= 0.05$, Shapiro-Wilk) and that the difference in linguistic features, including informational density, between students with higher and lower English proficiency levels is not statistically significant. Students with higher levels of English proficiency used more nouns ($P = 0.990 > 0.05$), verbs ($P = 0.750 > 0.05$), adjectives ($P = 0.994 > 0.05$), adverbs ($P = 0.998 > 0.05$), and pronouns ($P = 0.896 > 0.05$). Students with lower English proficiency levels appeared to show nouns ($P = 0.857 > 0.05$), verbs ($P = 0.923 > 0.05$), adjectives ($P = 0.875 > 0.05$), adverbs ($0.860 > 0.05$), and pronouns ($P = 0.992 > 0.05$).

The relationship between students’ level of proficiency and the level of informational density in their academic essays

The present study was designed to investigate the informational density of Indonesian graduate students enrolled in Hungarian higher education with vastly differing levels of English proficiency. The following measures of the use of informational density in student texts are depicted in Fig. 2. The frequency of content words is ascertained by informational density (i.e., nouns, verbs, adjectives, adverbs, and pronouns). The use of verbs had the highest difference in informational density between students with higher and lower English proficiency levels,
including the independent sample test analysis, followed by nouns, adverbs, pronouns, and adjectives. According to with statistical value, the difference is not significant.

Figure 3 depicts a comparative assessment of content words in texts from different academic disciplines written by students’ diverse levels of English proficiency. Nouns were the most common informational density type in texts written by Indonesian graduate students, followed by verbs, adjectives, adverbs, and pronouns. In their texts, students with higher and lower English proficiency used more nouns. On the other hand, students in higher-grade levels appeared to use more nouns than lower-grade students.

According to quantitative tests, there is a link between English proficiency levels and the use of informational density. The analysis confirmed that the different value between these two categories was not statistically significant despite those mentioned earlier. Figure 4 likens higher and lower levels of English proficiency in terms of informational density frequency number. It was revealed that students with higher English proficiency levels have a higher total informational density than students with lower English proficiency levels.

The relationship between students’ level of proficiency and the level of abstraction in their academic essays

It appears believed that nominalization is a fundamental aspect of academic writing, with abstraction being one example (Hyland, 2020). This research investigated the abstraction using

![Figure 2. The different value of informational density between C1 and B2 level students](image)

![Figure 3. The comparison of informational density between C1 and B2 level students](image)
the Coh-Metrix webtool analysis criteria, which included three main categories: nominal phrases, imagibility of content words, and concreteness. The findings revealed that texts written by Indonesians with higher and lower English proficiency levels contained more nominal phrases (see Fig. 5).

In addition, Fig. 6 shows that the frequency of abstraction was higher in the category of higher-level students than lower students. Texts written by students with low English proficiency are more concrete than texts written by students with higher English proficiency. As a result, students with higher levels of English produced more abstract texts than their classmates.

The findings indicated a link between Indonesians’ C1 and B2 English proficiency levels in terms of abstraction. Figure 6 demonstrates that abstraction is a key feature in academic writing. The more abstract the text is, the more difficult the text is. However, it is not the case, and the most important thing is that abstraction has closely related to high-level academic texts (Hood, 2010).
Informational density and abstraction are linguistic features that are only valid in academic writing. They assist in the comprehension of writing proficiency and the quality of writing (McNamara et al., 2010). The linguistic features that appeared in graduate student texts written by Indonesians in three different disciplines were analyzed and discussed. The study attempts to identify two questions: the relationship between students’ levels of proficiency and informational density in their academic essays and the other regarding the relationship between students’ levels of proficiency and the level of abstraction in their academic essays.

In terms of informational density, the overall findings of this study indicate that C1 level students produce texts with more content words. There is no statistically significant difference in informational density between students at the C1 and B2 levels. The analysis revealed that the texts of B2 students are more concrete than the texts of C1 students when abstraction is taken into account. As a result, C1 students’ texts are more abstract than B2 students. There is no statistically significant difference in imageability, concreteness, or nominal phrases between texts written by C1 and B2 students.

Whereas this work remains to be done to understand the linguistic features in academic writing better, this preliminary study attempted to shed some light on how informational density and abstractions are taken into account in academic texts written by Indonesian graduate students in three different fields. This present study faces some limitations and needs further action to provide better results, including the field of study and the sample size. Research is needed to compare academic writing registers in each discipline among students with varying levels of English proficiency. In terms of sample size, providing a larger sample size leads to accommodating significant results and brings the representativeness more closely to the population.

For all that, the present study adds valuable insights in terms of the new population and a new dimension to Hungarian higher education, giving prominence to both theoretical and practical perspectives. Commensurate with the theoretical framework, abstraction and informational density are key linguistic features to consider when evaluating academic writing quality, especially dealing with high-stakes texts. As a result, it is strongly advised for pedagogical instruction to address the features as soon as students begin their studies at the
university. The current study’s findings show that higher-level students produce more of those features than their counterparts. The results demonstrate that producing high-stakes academic writing is an important component of academic success in higher education.

Conflict of interest: The author declares no conflict of interest.

ABOUT THE AUTHOR

Dedy Subandowo is a Ph.D. student of Applied Linguistics in the Department of Theoretical Linguistics, Pázmány Péter Catholic University (PPCU), Hungary. His research interest involves L2 academic writing, academic discourse, and writing strategies. His doctoral research is about Indonesian graduate students’ English academic writing challenges in a non-native English-speaking environment.

ACKNOWLEDGEMENTS

I would like to thank the Tempus Public Foundation from the Hungarian Government for its Stipendium Hungaricum scholarship funding support and the anonymous reviewers for their detailed and thoughtful comments on earlier versions of this paper.

REFERENCES


