

Identifying Vocational Student's Familiarity with Vocabulary in Academic Context in Abstract Writing

Satriya Bayu Aji ¹, Eka Mutia Lubis ²

¹ Geomatics Engineering Study Program, Informatics Engineering Major,
Politeknik Negeri Batam, Batam, Indonesia

Email: satriya@polibatam.ac.id

² Electrical Engineering Study Program, Electrical Engineering Major,
Politeknik Negeri Batam, Batam, Indonesia

Email: mutia@polibatam.ac.id

Article Info

Article history:

Submitted Sept 08, 2020

Revised Sept 29, 2020

Accepted Nov 02, 2020

Published Nov 03, 2020

Keywords:

Abstract
Academic Writing
Parallel Corpus
NGSL
NAWL

ABSTRACT

In order to write an abstract that appeals to a broader audience, it is important to understand the vocabulary used in the context. By identifying words used in the abstracts of undergraduate thesis from Politeknik Negeri Batam students and comparing them with those used in other academic writings, this research attempts to show that developing student's awareness of academic writing genre is critical for their progress. The data were taken from the abstract section of undergraduate thesis of Politeknik Negeri Batam students published in 2018, which were compiled into a corpus. From the corpus, the keywords and word profile would be identified, which could then be compared to other corpora. Two word lists (New General Service List and New Academic Word List) were selected as the level lists in carrying out word profiling during the analysis. The results show that the students lacked the necessary words and had not yet recognized the genre of academic writing.

Corresponding Author:

Satriya Bayu Aji,

Geomatics Engineering Study Program, Informatics Engineering Major,
Politeknik Negeri Batam

Jl. Ahmad Yani, Batam Kota, Batam, Indonesia.

Email: satriya@polibatam.ac.id

INTRODUCTION

In order to more easily convince the readers, an abstract must sequentially include motivation/goals, problem formulation, methods/approaches, results, and conclusions (Koopman, 1997). Prior knowledge consists of direct and verbal experiences. These allow the discourse participants to evaluate the truth, appropriacy, and relevance of a proposition. In unfamiliar situations, different schemata (prior knowledge of rhetorical structures) must be activated. In this research, "A genre comprises a class of communicative events, the members of which share some set of communicative Purposes" (Swales, 1991: 58). Genres may vary according to their complexity, how much they are prepared beforehand, readership consideration, and universal/language-specific tendencies.

Academic Word List (AWL) is designed primarily for teachers as part of a program to prepare students for further study or to be used independently by students to learn words

that need to be understood at the tertiary level (Coxhead, 2000: 213–238). This list contains 570 words (word family), with entries (headword) in the form of the basic form (stem) of these words. Names of places, people, countries are not included in the list, nor are Latin forms such as *et al*, *etc*, *ie*, and *ibid*. The data comes from Academic Corpus, which contains more than 3,500,000 words in 414 texts. AWL requires mastery of vocabulary in General Service List (GSL).

GSL is a collection of 2,000 selected words most needed in learning English. GSL is a list of words intended to assist learners of English language. After the initial appearance of GSL in 1953, many other lists emerged. An attempt to improve this list was carried out by Bauman and Culligan (1995) by using standards proposed by Bauer and Nation (1993: 253–279). This improvement increased the number of words to 2,284. Another weakness of the early version of GSL is the size and age of the source texts.

NGSL was developed based on more than 273 million words from Cambridge English Corpus (CEC). NGSL offers an increase in coverage (Browne, 2013: 13). However, the number of words that must be studied to get a 1% increase in coverage increased sharply after reaching 92%. New Academic Word List (NAWL) was developed based on an academic text corpus containing 288 million words from Cambridge English Corpus (CEC), Michigan Corpus of Academic Spoken English (MICASE), and British Academic Spoken English (BASE). The coverage of NGSL and NAWL is 5% more than GSL and AWL (Browne, Culligan, & Phillips, 2013).

Researches related to word lists can be done by comparing the frequency of words occurring in a corpus comprising texts from several teaching materials with words in a word list. Based on the research by Astika and Kurniawan (2018: 630), which compared the frequency of occurrences of words in textbooks with New General Service List (NGSL), the total number of entries in the textbook is 58% of the total number of entries in NGSL. The data were obtained from the corpus that contains 147,199 words. As a follow-up, the researchers suggest to compile a dictionary based on the entry. This study illustrates how much the proportion of English vocabulary in NGSL is used by speakers with Bahasa Indonesia as their first language.

Another list that has been developed is the Science Academic Word List (SAWL), which consists of 432 words that are needed for students to understand natural science journal articles (It-ngam & Phoocharoensil, 2019: 657). SAWL has a wider coverage (5.82%) in the compiled corpus, compared to Science-specific Word List (Coxhead & Hirsh, 2007: 65–78). The organization of SAWL is based on the corpus of Scientific Academic Journal (SAJ), which consists of 5.5 million words. The corpus used in the compilation of SAWL consists of journal articles consisting of 1,062 articles in 11 disciplines.

Another word list is British National Corpus/Corpus of Contemporary American English (BNC/COCA) word list. The list is divided into 29 groups. The first 2,000 words in the BNC/COCA word list can be used as an alternative to GSL. These two groups (1st–1000th and 1001st–2000th word) were arranged according to a special corpus consisting of ten million tokens. Here, tokens are defined as sequences of letters separated by spaces or punctuations. Six million tokens are in the form of oral texts, as well as films and television shows, in British and American pronunciations. The written texts also include children's literature and fiction. This step was taken to avoid the tendency towards formal written variations. The criteria used in determining the words included in the list are based on the

classification according to Nation (2012) at level 1 to 6. BNC/COCA word list includes compound words but excludes phrases. There are 272,782 entries (word types) in BNC corpus that are not included in the first 20 groups, plus a number of proper nouns, transparent compounds, exclamations, markers of doubt, and other marginal words found in oral communications.

To compare GSL, NGSL, and BNC/COCA word list, Kwary and Jurianto (2017: 60–72) built a corpus with the data derived from five news articles displayed on MTV Asia website on April 1st, 2015. This research reminds us that the meaning of a word changes from time to time and tends to differ from one context to another. The compilation of those word lists serves different purposes.

As a part of an undergraduate thesis, an abstract in the context of Indonesian undergraduate thesis is interesting due to the need to include two abstracts: one in Bahasa Indonesia and the other in English. Therefore, it is interesting to compare the English-translated abstracts to other English academic writings. The previous paragraphs illustrate that a corpus can be beneficial in providing the frequency count of words. Using methods in corpus linguistics, this research aims to compare their word profile. The word profile is the basis on which the claim regarding potential discrepancies between the translated abstracts and English language academic writings in terms of vocabulary is made.

THEORY AND METHODS

To compare the word profile of the English translation of the abstracts with the reference corpora, a corpus was compiled. The abstract section of undergraduate theses of Politeknik Negeri Batam students published in 2018 was used as the data in this research. They consist of 496 original texts in Bahasa Indonesia and their English translation. All abstracts were taken from the website <https://repository.polibatam.ac.id>. Not all abstracts from every study program in Politeknik Negeri Batam were included as the data due to the limitation in the search and filter feature of the website. In the source text, there are 78,312 word tokens, comprising 6,937 word types. In the target text, there are 87,003 word tokens, comprising 6,657 word types. The heading *Abstrak* and *Abstract* in the source and target text are not part of the data. This research also made an assumption that a title is not part of an abstract. Therefore, it was set aside from the analysis. Based on the field of study, the abstracts were arranged into eight different categories, which are shown in Table 1.

Number of Abstracts	
Field of Study	Number of Abstracts
Business Administration	51
Accounting	86
Managerial Accounting	58
Electrical Engineering	60
Informatics Engineering	68
Multimedia and Network Engineering	57
Mechatronics Engineering	74
Mechanical Engineering	42
Total	496

The collected data must first be put into the format compatible with the concordancer software. The original .pdf files were first converted into .txt format in UTF-8 encoding. The abstracts in the source and target language were then aligned to make sure that they were displayed correctly in the parallel concordancer. In this research, AntConc 3.5.88 was used as the concordancer. The concordancer is useful in generating word lists, frequency counts, and keywords. For the parallel concordancer, AntPConc 1.2.19 was used to provide the keyword in context (KWIC) display for both the source and target text simultaneously.

A concordancer helps during the search for a word or phrase in a corpus. Several things can be done utilizing a concordancer, such as investigating the central and typical meaning, meaning distinctions, as well as meaning and its pattern. However, a concordancer only displays information relevant for an inquiry. How the information is interpreted depends on the perspective of the user. One type of corpus interpretation involves distinguishing between categories, another involves making generalizations about the association between the way a word is used and its meaning (Hunston, 2002: 39–65).

In conducting vocabulary profiling, AntWordProfiler 1.4.110 was used. The level list was based on NGSL and NAWL. The first 1,000 words in the NGSL made up the 1st NGSL category. The rest constituted the 2nd NGSL category. Moving up the category, there was NAWL. Words not included in either NGSL or NAWL were grouped under the \emptyset category. AntConc uses the term keyness value to refer to the value generated by the software to measure the strength of a keyword in a corpus. To measure the keyness value, Log-Likelihood was used as the Statistical Measure. The cutoff point was set to $p < 0.05$. For the Effect Size Measure setting, the Dice coefficient was selected. The same settings were also applied to the two reference corpora. Brown Corpus represents American English and Lancaster-Oslo-Bergen (LOB) Corpus British English. Only the learned and scientific writings section of both reference corpora were used during the comparison. The two corpora were chosen due to their comparable size and composition as well as the consideration of availability despite the age and size limitation of these corpora.

RESULT AND DISCUSSION

From the compiled corpus, a word list was generated for each field of study. A keyword list was then generated from each word list. Ten words with the highest keyness value for the eight fields would be the focus of this research. The list of these top ten word tokens is shown in Table 2.

Top 10 Keywords

Field	Source	Target
Business Administration	<i>responden, fisik, berpengaruh, kinerja, kepuasan, variabel, penelitian, analisis, signifikan, faktor</i>	employee, physical, sampling, respondents, factors, satisfaction, significant, performance, study, analysis
Accounting	<i>internal, pengendalian, membahas, kredit, tugas, akhir, kas, pajak, pt, prosedur</i>	final, inventory, credit, words, internal, cash, tax, procedures, company, pt, procedure
Managerial Accounting	<i>corporate, diukur, periode, kinerja, keuangan, pengaruh, berpengaruh</i>	exchange, performance, stock, samples, capital, effect, research, financial,

Field	Source	Target
Electrical Eng.	<i>variabel, sampel, perusahaan, penelitian power, mobil, listrik, kabel, alat, arus, sensor, kentang, tegangan, motor</i>	corporate, study plant, supply, module, sensor, current, potato, power, motor, voltage, control
Informatics Eng.	<i>php, twitter, kartu, log, event, surat, berbasis, web, informasi, aplikasi</i>	meeting, paper, internship, location, php, based, twitter, web, information, application
Multimedia and Network Eng.	<i>interaktif, rendering, multimedia, anak, render, media, game, video, animasi, film</i>	interactive, development, graphic, film, multimedia, media, game, video, animation, rendering
Mechatronics Eng.	<i>gawang, error, warna, mendeteksi, sensor, manusia, pergerakan, sistem, bola, robot</i>	image, camera, movement, error, sensor, system, detection, color, ball, robot
Mechanical Eng. Total	<i>steel, korosi, mm, proses, nozzle, mesin, material, kekasaran, permukaan, kapal batam, dilakukan, proses, digunakan, sistem, metode, perusahaan, hasil, data, menggunakan, penelitian</i>	process, steel, machine, material, mm, nozzle, plate, roughness, surface, ship batam, results, analysis, process, method, system, used, study, research, data, using

Even though most rows contain exactly 10 words, 11 words were included in some fields. In the Managerial Accounting source text keyword list, both *berpengaruh* and *pengaruh* were included. Some would argue that this inclusion is unnecessary since the word *berpengaruh* can be derived from the word *pengaruh* using the derivational affix *ber-*. However, a derivational process will result in a word with a different part of speech. From figure 1, it can be observed that the word *berpengaruh* is used as a verb, as in the line *...pengujian hipotesis berganda penelitian ini terdapat tiga variabel yang berpengaruh negatif dan signifikan terhadap net profit margin, yaitu working...* In this line, *berpengaruh* functions as the predicate of the clause that necessitates the verb *berpengaruh*. On the other hand, the word *pengaruh* is used as a noun. For example, in the line *Profitabilitas memiliki pengaruh yang kuat terhadap hubungan antara ISO 14001 dan nilai perusahaan*, the word *pengaruh*, which is a noun, functions as the direct object.

Line	KWIC
14	...jika pelajaran, media yang digunakan dirasakan kurang menarik, dan hal tersebut berpengaruh kepada kurang pemahamannya mahasiswa terhadap materi yang di
15	...sikap berpengaruh positif dan signifikan terhadap keberhasilan usaha, 2) motivasi berpengaruh negatif dan signifikan terhadap keberhasilan usaha, 3) kemampuan t
16	...nifikan terhadap kepuasan pengunjung sedangkan variabel jaminan (assurance) berpengaruh negatif dan signifikan terhadap kepuasan pengunjung.
17	Hasil penelitian menunjukkan bahwa secara parsial variabel kepemimpinan berpengaruh negatif dan tidak signifikan terhadap kinerja karyawan, dan variabel
18	...postif dan tidak signifikan terhadap minat berkunjung, sedangkan variabel orang berpengaruh negatif dan tidak signifikan terhadap minat berkunjung.
19	...ari pengujian hipotesis regresi berganda penelitian ini terdapat tiga variabel yang berpengaruh negatif dan signifikan terhadap net profit margin, yaitu working capi
20	Hasil penelitian ini antara lain yang pertama, komisaris wanita berpengaruh negatif terhadap kinerja perusahaan.
21	Kelima, latar belakang pendidikan direksi berpengaruh negatif terhadap kinerja perusahaan.
22	Sedangkan diversifikasi produk berpengaruh negatif terhadap profitabilitas perusahaan.
23	...bahan penjualan tidak berpengaruh terhadap struktur modal, tetapi profitabilitas berpengaruh negatif terhadap struktur modal dan secara simultan ukuran perusal
24	Keahlian Komite Audit signifikan berpengaruh negatif terhadap manajemen laba dengan menggunakan discretionary
25	...n menunjukkan bahwa struktur modal yang diukur dengan debt to equity ratio berpengaruh negatif terhadap return of equity, dan return of asset
26	Reference
16	This research uses multiple regression, the result of this research which show the variable of reliability, responsiveness, empathy, and tangible have positive and significant to customer satisfaction while variable assurance have negative and significant to customer satisfaction ...
17	The results showed that the leadership of the influential variables are partially negative and insignificant against the performance of employees, and work motivation variables positive and significant effect on performance of employees.
18	Partially, product and promotion are affecting positively and significantly in interest visiting Love Seafood Restaurant, Tanjung Piyau. Price, place, process and physical evidence are affecting positively and not significantly in interest visiting Love Seafood Restaurant, Tanjung Piyau. And people is affecting negative and not significantly in interest visiting Love Seafood Restaurant, Tanjung Piyau.
19	Partially working capital turnover, days of supply and cash conversion cycle is significant negative effect on the net profit margin.
20	The results of this study include the first, female commissioners have negative value on the company's performance.
21	Fifth, the educational background is called negative towards company performance.
22	While the product diversification effect negatively to the company's profitability.
23	The results of this study indicate that size and sales growth has an effect on capital structure but profitability have effect negative on capital structure and simultaneously size, profitability and sales growth have effect on capital structure.
24	Audit Committee Member's Expertise has significant influence negative to earning management by use of discretionary accruals Jones's modification.

Figure 1 The KWIC Display of *berpengaruh*

Furthermore, the decision to include both *berpengaruh* and *pengaruh* in the keyword list also stemmed from the words chosen by the translator as their equivalents in the target text. From figure 1, it can be observed that the translators had five different choices in deciding the equivalent of the word *berpengaruh* in English. These words include *is affecting* in line 18, *effect* in line 19, *have ... value* in line 20, *is called* in line 21, and *has ... influence* in line 24. Some of the examples are indeed ungrammatical, for example, *While the product diversification effect(sic!) negatively to the company's profitability*. Here, the word *effect* (a noun) is incompatible with the function required by the clause (a verb). However, an extensive analysis for the entire corpora requires tagged corpus in Bahasa Indonesia.

This discrepancy between the function of a phrase in a clause with the part of speech does not happen in the example regarding the word *pengaruh*. This can be seen in figure 2. There are three words that were chosen by the translators as the English equivalent of the word *pengaruh*. These words include *effect*, *influence*, and *impact*. All of these words are nouns, which are the same part of speech as the word in the source text.

Line	KWIC
135	Penelitian ini bertujuan untuk menganalisis pengaruh utilitarian value dan hedonic value terhadap repurchase intention pada
136	Penelitian ini bertujuan untuk menguji pengaruh variabel independen yaitu reliabilitas (reliability), daya tanggap (respons
137	an untuk mengetahui tingkat kekasaran permukaan hasil proses milling terhadap pengaruh variasi komposisi coolant.
138	ikaan yang rendah(halus),sehingga menimbulkan permasalahan yaitu bagaimana pengaruh variasi komposisi cairan pendingin terhadap kekasaran permukaan mab
139	Profitabilitas memiliki pengaruh yang kuat terhadap hubungan antara ISO 14001 dan nilai perusahaan.
140	terja saham akan tetapi dengan adanya penerapan sertifikasi ISO 14001 memiliki pengaruh yang positif terhadap kinerja keuangan yang diukur oleh ROE dan ROA
141	Motivasi sebagai variabel moderasi mempunyai pengaruh yang positif terhadap kedua variabel ini, semakin besar motivasi yang d
142	Perusahaan sebagai variabel moderasi Good Corporate Governance mempunyai pengaruh yang positif.
143	Hasil penelitian ini menunjukkan secara parsial tidak semua variabel mempunyai pengaruh yang signifikan terhadap keputusan siswa dalam memilih SMK Teladan
144	Ketiga faktor tersebut memiliki pengaruh yang signifikan terhadap return on asset, sedangkan pendidikan dewan
145	hasilan Perusahaan, Profitabilitas, Struktur Asset dan Ukuran Perusahaan memiliki pengaruh yang signifikan terhadap struktur modal.
Line	Reference
136	The objectives of this research are: reliability, responsiveness, assurance, empathy and tangible of customer satisfaction.
137	Therefore, a comparative study is needed to find out the surface roughness of the milling process on the effect of coolant composition variation.
138	So required the composition of the appropriate coolant to obtain a low surface roughness value (fine), causing problems that is how the effect of coolant liquid composition variation against surface roughness of S45C carbon steel material On Turning Process.
139	Profitability has a strong effect on the relationship between ISO 14001 and the value of the firm.
140	The results of this study indicate that partially applying ISO 9001 certification does not prove a positive and significant influence on financial performance and stock performance but with the application of ISO 14001 certification has a positive influence on performance finance measured by ROE and ROA.
141	Motivation as a moderating variable has a positive influence on both of these variables, the greater motivation of the employee the better the level of performance given by an employee.
142	Firm Size as a moderate variabel has a positive impact.
143	The results of this study show that not all variables have significant influence on student decisions in choosing SMK Teladan Batam.
144	These three factors have a significant effect on return on assets, while the board of commissioners education and the board of director's education have no significant effect on return on assets.
145	The result showed that corporate income tax, profitability, asset structure and firm size have a significant influence on capital structure.

Figure 2 The KWIC Display of *pengaruh*

Here, there is no shift in the part of speech from the word in the source text to the word in the target text. From the figure, it can also be observed that, in the target text, the translators used *using* and *used* as the equivalents of the words *menggunakan* and *digunakan*, respectively, in the source text. Therefore, these two words were also included in the keyword list under the Total heading, which subsumes the eleven words with the highest keyness value in the eight compiled corpora combined.

However, the decision to include both the words *procedure* and *procedures* in the Accounting field might be considered problematic. Unlike the previous examples, which involve a derivational process, the affixation involved in the formation of the word *procedures* is inflection. Another problem was to decide which token should be ignored or whether the two tokens should be regarded as one. The decision here to regard those tokens as separate was done *ad hoc* as there was no hard and fast rule to deal with these

problems and to make up for the limitation imposed by the small size of the corpora compiled for this study so that more possible usage of those words could be taken into account. Consequently, this has some effects on the conclusion drawn from this research.

Table 2 shows that there are differences in the makeup between the source and target text for every field of study. Most lists in the target text contain only around six to eight literal renditions of the source text chosen by the translators. Choices made by the translators in a field of study could also affect the resulting keyword lists in other fields of study that were part of the compiled corpus. This change could be the result of standardization. Many words used in the target text fall in the first category of NGSL. This frequent use of words from the first category of NGSL can be observed in all eight fields of study. The detail is shown in Table 3. The table also shows a small percentage of word types used compared to the percentage of word token. This suggests that the translators only made use of a limited number of lexemes. The finding further strengthens the claim of standardization as one of the universals of translation (Hatim & Munday, 2004: 13).

The Student Corpora Word Profile

List	Token (%)	Type (%)	Group (%)
1 st NGSL	62669 (72.03)	1757 (26.39)	890 (16.91)
2 nd NGSL	7670 (8.82)	999 (15.01)	686 (13.03)
3 rd NGSL	3061 (3.52)	525 (7.89)	398 (7.56)
NAWL	2758 (3.17)	448 (6.73)	361 (6.86)
∅	10845 (12.47)	2928 (43.98)	2928 (55.63)
Total	87003 (100)	6657 (100)	5263 (100)

The frequent use of a limited number of word types surely affected the keyness value in general, not only these words. Less variation in vocabulary can make the reading easier because the words used are in the first category of NGSL and these are words with a high frequency that appears in many genres. However, the limited vocabulary used by the translators in the target text may also suggests that they had not yet understood enough vocabulary needed to write or translate a text in the academic genre. Here, the latter interpretation is more likely because the heavy use of words belonging to the first category of NGSL is not typical of the established English language academic writing in general. This corresponds to the finding regarding the moderate percentage of common vocabulary in textbooks mentioned in the introduction. The data in Table 4 shows this frequent use of words in the first category of NGSL in detail for the eight fields of study.

Word Profile (1st NGSL)

Field	Token (%)	Type (%)	Group (%)
Business Administration	6168 (70.6)	554 (44.43)	423 (40.13)
Accounting	9238 (73.18)	816 (43.75)	545 (36.07)
Managerial Accounting	6875 (72.28)	514 (46.98)	379 (41.51)
Electrical Engineering	7100 (69.74)	739 (41.68)	511 (34.88)
Informatics Engineering	8665 (75.33)	908 (45.08)	607 (38.03)
Multimedia and Network Engineering	9241 (72.88)	945 (42.59)	624 (35.02)
Mechatronics Engineering	10064 (71.16)	898 (41.63)	595 (33.83)
Mechanical Engineering	5318 (69.73)	620 (42.03)	440 (35.74)

Some words used in the target text are not part of either NGSL or NAWL. The percentage of those words for each field of study is shown in detail in Table 5. Words

included in this list consist of borrowed and technical terms as well as words that were mistyped. The compilation of NGSL and NAWL has coverage as its utmost priority. The aim is to cover as many genres as possible. Therefore, a highly specific and technical term as well as words that are circulating only in a specialized genre (for example, *nonparametric*, *hereditary*, and *actionscript*) are less likely to be included in either NGSL or NAWL. In borrowing, a term in the source language is used in the target text, either in its pure or naturalized form. Other than technical terms, external culture-specific references, such as the name of places (*Jakarta*, *Batam*, and *Kabil*), tribe (*melayu*), or vegetation (*meranti*), are also the prime candidate for borrowings. The last constituent of the words that are not included in both NGSL and NAWL is mistyped words. Mistyping should have been easy to spot and must be avoided at all cost, especially in writing an abstract. Mistyping comes in many forms. Minor problems may include errors such as **methode*, **humam*, and **resustance*. Even though these words can possibly hamper the reading process, they still retain some resemblance with their correct written form and their meaning can still be inferred from the overall context or surrounding words. On the other spectrum, there are major mistakes that will make a reading very hard, if not impossible. Examples of these problematic instances include **thepurposeofthisprojectisto*, **thatresultedinthemotorisnot*, and **efficiencyofthemotorwillbethesameasthefactorydefault*. This, in part, is the consequence of adopting the definition of word as a string of characters separated by spaces or punctuations. Every decision made by the translators in dealing with all these three categories (technical terms, external culture-specific references, mistyped words) may contribute to the differences in the makeup of the keyword list.

Word Profile (Level Ø)

Field	Token (%)	Type (%)	Group (%)
Business Administration	1017 (11.64)	305 (24.46)	305 (28.94)
Accounting	1666 (13.2)	554 (29.71)	554 (36.66)
Managerial Accounting	989 (10.4)	279 (25.5)	279 (30.56)
Electrical Engineering	1484 (14.58)	509 (28.71)	509 (34.74)
Informatics Engineering	1274 (11.08)	481 (23.88)	481 (30.14)
Multimedia and Network Engineering	1571 (12.39)	604 (27.22)	604 (33.89)
Mechatronics Engineering	1756 (12.42)	605 (28.05)	605 (34.39)
Mechanical Engineering	1088 (14.27)	418 (28.34)	418 (33.96)

Other than introducing difficulties in reading, the incompatibility of vocabulary profile with the existing English language academic genre may also lead readers to misinterpret the meaning of the texts. Both the authors and the consumers of these texts—not to mention the translators, which usually are the authors themselves—clearly do not want this problem to occur. Moreover, writings that do not adhere to the restrictions imposed by the existing genre will face resistances. Although a deviation can also mean a challenge to the established norm (to demand a change), it is unlikely that these are the changes preferred by the academic genre writing in Indonesia.

The incompatibility with the English language academic genre can be identified from the data in Table 6. In the table, the word profile of the compiled corpus is compared with the word profile of Brown Corpus and LOB Corpus. As can be seen from table 6, the comparison was only made for the word tokens, considering the difference that might present in how each corpus defined what counts as a word.

Comparison with the Reference Corpus

List	Student (%)	Brown (%)	LOB (%)
1 st NGSL	62669 (72.03)	2366 (17.03)	2280 (16.71)
2 nd NGSL	7670 (8.82)	1757 (12.65)	1681 (12.32)
3 rd NGSL	3061 (3.52)	1075 (7.74)	1031 (7.56)
NAWL	2758 (3.17)	1055 (7.59)	980 (7.18)
∅	10845 (12.47)	7641 (54.99)	7674 (56.24)

The level lists that consist of NGSL and NAWL are considered sufficient since other word lists might have differences in their design to serve their purpose and the difficulty to compare them, as has been stated in the introduction. These data also strengthen the previous claim that the translators of these abstracts had not yet understood enough vocabulary needed to write or translate a text in the academic genre.

A close resemblance can be observed for both Brown Corpus and LOB Corpus in terms of their word profile. By comparing the compiled corpus with the two reference corpora, some differences can be identified. Most words used in the texts constituting the two corpus sections are words that did not make their way into either NGSL or NAWL. The use of specialized terms and jargon that are widespread in the academic genre writings is the norm that can be observed here. These highly technical terms facilitate participants in the discourse to achieve a high level of precision in formulating definitions, not to mention their exclusive nature. Due to the specific meaning and restricted use, it can be understood why these words were not included in either NGSL or NAWL, which was designed for different purposes.

CONCLUSION

It can be concluded from the results and discussions above that the students had not yet mastered enough words in their vocabulary. This claim is based on the heavy use of the first category of NGSL and a relatively low percentage in the use of words outside NGSL and NAWL. Some might argue that these differences might not be caused by shifts in the target text but, instead, resulted from the different conventions of the academic writing exercised in the source texts, with a high level of repetition of only a few familiar words and limited use of specialized terms and jargons. However, such counter-claim must be based on solid evidence, ideally from a sizable, balanced, and representative example of texts from a diverse set of genres in the source language, which, at the time of writing this article, had not yet existed.

What is clear from these differences is the negligence of the English language academic writing genre in the abstract section by the translators. Indeed, the articles from where these abstracts come from had undergone some revisions. However, this raises further doubt about the capacity of the advisors, especially regarding the English language expertise needed in this genre. The occurrence of such a huge discrepancy with the accepted norm requires immediate attention and change in policy regarding undergraduate thesis writing and the role of English language teachers in this institution if it is true that the dissemination of information is the real goal of academic writing. The awareness regarding the importance of language in shaping discourse needs to be raised,

not only for students but also for everyone who is part of the discourse community. From the findings, it is hard to resist the impression that the inclusion of the English translation of an abstract is merely to fulfill the requirement dictated by the guideline.

ACKNOWLEDGEMENT

The authors of this article would like to offer their thanks to Politeknik Negeri Batam, especially Pusat P2M (Center for Research and Community Service) for the financial support and KKT Bahasa dan Budaya (Applied Research Group for Language and Culture) for the administrative support.

REFERENCES

- Astika, G., & Kurniawan A. (2018). Lemmatizing textbook corpus for learner dictionary of basic vocabulary. *Indonesian Journal of Applied Linguistics*, 7(3), 630–637.
- Bauer, J., & Nation, I.S.P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279.
- Bauman, J., & Culligan, B. (1995). About the general service list. Cited on April 12, 2020 from John Bauman's website: <http://jbauman.com/aboutgsl.html>
- Browne, C. (2013). The New general service list: celebrating 60 years of vocabulary learning. *The Language Teacher*, 37(4), 13–16.
- Browne, C., Culligan, B., & Phillips, J. (2013). The new academic word list. Cited on April 12, 2020 from New General Service List website: <http://www.newgeneralservicelist.org/nawl-new-academic-word-list>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Coxhead, A., & Hirsh, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, 12(2), 65–78.
- Hatim, B., & Munday, J. (2004). *Translation: an advanced resource book*. London: Routledge.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Nation, I.S.P. (2012). The BNC/COCA word family lists. Cited on April 12 2020 from Paul Nation of Victoria University of Wellington's bibliography page: https://www.wgtn.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf
- It-ngam, T., & Phoocharoensil, S. (2019). The development of science academic word list. *Indonesian Journal of Applied Linguistics*, 8(3), 657–667.
- Koopman, P. (1997). How to write an abstract. Cited on April 12, 2020 from Philip Koopman of Carnegie Mellon University's website: <https://users.ece.cmu.edu/~koopman/essays/abstract.html>
- Kwary, D. A., Jurianto. (2017). Selecting and creating a word list for English language teaching. *Teaching English with Technology*, 17(1), 60–72.
- Swales, J.M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.