



Lexical Richness of One-Paragraph Comments on Facebook by University Students in ‘Purposive Communication’ Course

Leonardo O. Munalim

*Dean, School of Arts and Sciences, Philippine Women’s University (PWU), Manila,
Philippines*

Biodata: **Dr. Munalim** obtained a PhD in Applied Linguistics from Philippine Normal University; finished a certificate in distance education from University of the Philippines Open University; and is associated with *Journal of English as an International Language*, *Asian ESP Journal*, *TESOL International Journal*, *IAFOR Journals*, and other SCOPUS-indexed journals. He is the Research Director at PWU.

Email: lomunalim@pwu.edu.ph

Abstract

This exploratory study aims to investigate the lexical richness of one-paragraph comments on articles posted on students’ Facebook groups. It specifically measures the lexical sophistication, density and diversity of these comments. Variations are accounted for based on students’ disciplines such as Tourism and Marketing. A total of 23,502 tokens produced by 36 university students taking *Purposive Communication (in English)* for the academic

year 2019-2020 formed the corpus of the study. D_Tools, AntWordProfiler, TagAnt and UAM corpus tools were used. Results reveal that the overall one-paragraph comments are lexically sophisticated (84.52%); lexically poor in academic and technical words (3.93%); lexically dense (57.84%); and lexically diverse with the D value of 107.199. The Marketing cluster's levels of sophistication, density and diversity are statistically higher, except for the lexical item of nouns. Implications for ESP teaching-learning, and paragraph quality vs. quantity are offered. Limitations and recommendations are also offered toward the end of the paper.

Keywords: lexical density, lexical diversity, lexical richness, lexical sophistication, Facebook comments, ESP.

1. Introduction

Lexical level offers pragmatic roles in assessing proficiency, diagnosing progress and testing learners' language development, exposure and performance (cf. Schmitt, 2010; Szudarski, 2018). Consequently, within the remit of English for Specific Purposes (ESP), learners' spoken and written vocabulary, especially on lexical richness–diversity, sophistication and density– has become a burgeoning research enterprise (cf. Adolphs & Schmitt, 2003; Cooper, 1993; Daller, Van Hout, & Treffers-Daller, 2003; Dang & Webb, 2014; Kyle & Crossley, 2015). Johansson (2008) examined the patterns of lexical diversity in narrative and expository written genres. She found out that lexical diversity is a better measure to detect differences between age groups. Coxhead and Demecheleer (2018) investigated the technical vocabulary of plumbing, and found out that the plumbing written texts are lexically closer to written academic texts than to spoken texts. Other scholars (e.g. Chung & Nation, 2004; Hsu, 2014; Todd, 2017) also investigated technical words in anatomy and engineering disciplines. Higginbotham and Reid (2018) studied the lexical sophistication of L2 learner's academic essays, and found out that there is a significant relationship between the types of words produced and the written proficiency.

With these studies in mind, the mission of my study is to further engage in this vibrant research enterprise with regard to the learners' lexical sophistication, density and diversity in particular, and in learners' vocabulary in general. Meanwhile, it is understood that many factors affect the acquisition of lexical skills in L2. Vedder and Benigno (2016) mention a few of the factors such as "transfer from L1 or other languages, word length and pronounceability, explicit versus implicit word learning, vocabulary knowledge in relations to other linguistic skills, or the relationship between the quantity and quality of word knowledge" (p. 25). This present exploratory study, however, is not an attempt to treat these factors, and should be left for future inquiries. Rather, this present study only attempted to compare the lexical richness of two groups of students from the two disciplines in one-paragraph comments, and proves whether this controlled type of writing could still generate lexical richness.

My study is still moored in lexical richness, but the type of data under investigation is arguably innovative. It is composed of one-paragraph comments posted on our Facebook groups. To my knowledge, this type of paragraph-restricted comments has not been studied locally and internationally. The Filipino students' compositions are oftentimes digressive. This axiomatic claim is supported by studies in contrastive rhetoric, maintaining the Filipino's spiral and indirect discursive pattern (cf. Kaplan, 1966). Kaplan's assertion is supported by other studies, pinpointing that this roundabout pattern is actually induced by high-context communication style among Filipinos, which can result in circumlocutive responses (Clyne, 2002; Enriquez, 1992 as cited in Pe-Pua & Protacio-Marcelino, 2000). Therefore, Facebook comments restricted to one paragraph are served with the best intention to train my ESL students to identify key points and to prevent them from posting rambling answers.

Restricted one-paragraph comments are warranted by the perceived labor of reading and writing/commenting. To downplay the possible higher affective filter (Krashen, 1982), it was a prudent decision of mine to restrict the comments in one paragraph, in not more than five sentences. It is in this pragmatic goal in mind that I conceptualized this controlled type

of writing to help teachers analyze whether students' paragraphs are lexically rich, given the need for them to be succinct. Vocabulary or lexical richness is essentially indicative of their ability to compose their thoughts under paragraph-limit constraints.

The merit of my innovative use of restricted one-paragraph comments may be approved by Hutchinson and Waters (1991). Accordingly, the choice of test (activities and tasks) in ESP is not asking whether the text is authentic, but asking what role teachers would want the text to play in the learning process. I argue that the texts under analysis pass and qualify for the "practical concept of 'fitness to the learning purpose'" (Hutchinson & Waters, 1991, p. 159). However, my imposed one-paragraph comments may not be new at all. Standardized testing examinations such as IELTS writing portion has a number of words restrictions, for example, at least 150 words or 250 words (cf. Cambridge University Press, 2005). Looking at lexical richness, Smith and Kelly (2002) in their study of classical literature also restricted the analysis of the plays into 300-word segments. Higginbotham and Reid (2018) imposed a 2000-word essay on their study. Recently, Thawarom and Singhasiri (2020) investigated lexical richness in a one-minute speaking task.

The corpus of one-paragraph comments makes this present study distinct. Within the sphere of ESP, the learner corpus was divided into disciplinary groups such as Tourism and Marketing to see if variations in lexical richness can be accounted for when comments were grouped by academic disciplines. Thus, this present exploratory study using a corpus of one-paragraph Facebook comments is an attempt to answer these questions:

1. What is the lexical richness of one-paragraph comments on articles posted on Facebook in terms of sophistication, density and diversity?
2. Does lexical richness in terms of sophistication, density and diversity differ between the disciplines of Tourism and Marketing?

2. Literature Review

2.1. ESP and Vocabulary

For the last forty-four years, English for Specific Purposes (ESP) has been shaped through theoretical influences of (applied) linguistics, sociolinguistics, communicative language teaching, writing across the curriculum, learning theory and genre studies (Benesch, 2001). ESP refers to the teaching-learning of English as a second and foreign language, where the activities, tasks, and the overall curriculum is curated for real-life specific domains. These domains include, but are not limited to, business, economy, politics, technology, tourism, commerce, education, journalism, hotel management, and other genre-based forms of communication. ESP focuses on specific language skills as demanded by specific activities or professions, where English is needed to carry out specific goals (Benesch, 2001; Brown, 2016; Dudley-Evans & St. John, 1998; Paltridge & Starfield, 2012). Classroom activities should gear toward the development of students' specific skills to help them prepare for the real-world. One of these skills to be nurtured is the writing skill.

In writing, vocabulary has the special status in L2 learning (Alderson, 2005) in general and in ESP in particular. In fact, Dudley-Evans and St John (1998) maintain that different types of vocabulary are used in general life contexts, both in scientific and technical descriptions and discussions. Vedder and Benigno (2016) stress that lexical richness is part and parcel of the indicators of lexical proficiency. And this lexical proficiency and the “well-used vocabulary are likely to have a positive effect on the reader” (Laufer & Nation, 1995, p. 307). As Carter (2012) crisply puts it, “lexical items are powerful indices of ‘expression’” (p. 178). In this present study, the choice of vocabulary indicates the writers' ability to communicate effectively amid the word-limit constraints. Therefore, vocabulary regardless of its level of sophistication should be learned, taught, and researched in a language classroom.

The data of this present study are the comments on articles posted on students' Facebook groups. Amid the continuous breadth of social networking sites (cf. Christakis & Fowler, 2011; Clark, Hergenrader & Rein, 2015), the students should be able to write with quality in order to elevate healthy discourses on social media. I argue that there is a perceived need for understanding the students' lexical richness if teachers' intention is to help combat lying, misleadingness, esoteric truths, creative truths and agnotology (Block, 2019), in short, fake news, in media discourses. These harrowing attributes may be mitigated if our students are able to write purposefully and responsibly, without being too verbose in their comments.

Although vocabulary is one factor that impedes students from developing academic literacies (Johns, 1997), its importance in ESP cannot be underestimated. My study, which makes use of paragraph-restricted comments, may sit well with Crawford Camiciottoli's (2007) idea of "interpersonal and conventional aspects of texts, as it provides greater insight into the nature of academic discourse which involves differing social relationships among participants," (pp. 15-16) including me as the professor of the course. Overall, students' lexical richness may help them become ready in the real world, thereby inducting themselves into professional and healthy discourse communities in and out of the brick-and-mortar classrooms. As a matter of fact, it is through teaching and learning vocabulary in an ESP class that the students should be able to feel that they belong to a particular group, and eventually "become fully-fledged members of a particular community (Coxhead, 2013, p. 116).

2.2. Lexical Richness and Media Literacy

The increasing use of English has been propelled by the use of different social media platforms. Discourses in social media have become ubiquitous. Not only have these social media had a profound impact on students' discourses, but they are also being increasingly used to weaponize a "war of words" with the concept of free speech in mind. In their

attempts to get heard at the wider sphere, they use vocabulary – both sophisticated or not. Situating from the concept of genre as a social action in ESP, Paltridge (2013) claims that “through our use of genres, we say who we are and how we wish to be seen... we present ourselves to the world and take on (or reproduce) particular identities” (p. 354). In this present study, these word-limited comments may echo what Bloch (2013) has asserted in his article on “Technology and ESP” that technology has become the “space for creating new forms of communicating” (p. 387). Likewise, Starfield (2013) has mentioned the critical and pragmatic perspectives on ESP. As an extension to this notion, my paper takes up the same line of thinking that lexical richness or poorness in a controlled number of words may hasten the need for “multiliteracies” or “multiple literacies” in the epoch of the “most dramatic technological revolutions in history” (Kellner, 2000, p. 21). In short, lexical richness may have something to do with the need for media literacy, hence should be examined in word-limit comments to further help the ESP teachers to examine lexical proficiency and literacy in computer-mediated communications like Facebook comments.

2.3. Lexical Richness

Vocabulary richness has been rife with “diffusions and confusions” (Yu, 2009, p. 238). Many terms have been used in measuring a range of vocabulary in a text or conversation, such as lexical diversity, lexical range, vocabulary richness, flexibility and verbal creativity (cf. Malvern & Richards (1997). Lexical diversity and vocabulary richness may be also synonymous (Malvern, Richards, Chipere, & Duran, 2004).

For this present study, I favor the use of “lexical richness” by Read (2000) for consistency. He enumerates that lexical richness is measured through (1) type-token ratio or lexical variation, (2) lexical sophistication, (3) lexical density and (4) number of errors. I excluded the last measure with the idea in mind that the current paper is not for “error analysis.” Nonetheless, the first three measures are intuitively rich to provide preliminary insights of students’ lexical richness in paragraph-restricted article comments. Meanwhile, for the

type-token ratio or lexical variation, I follow Malvern and Richards's (1997) term "lexical diversity" for consistency. Thus:

2.3.1. Lexical Sophistication:

According to Read (2000), lexical sophistication is a "selection of low-frequency words that are appropriate to the topic and style of the writing, rather than just general, everyday vocabulary. This aspect of the range of expression also includes the use of technical terms and jargon..." (p. 200). Technicality of words also resides in the collocation of frequent words. Likewise, Milton and Hales (1997) stressed that the use of infrequent vocabulary may be a feature of a technical language. Using a sample of a Nissan technical manual, they reported that the use of long and repetition of noun phrases can be an indication of technical expressions.

Beck, McKeown, and Kucan (2002, 2005) bare three tiers of vocabulary: Tier 1 is composed of the most familiar words of 8,000 word families known to an average of third grader; Tier 2 is where the academic success can occur because this tier is composed of 7,000 word families, which are not limited to one content area; and Tier 3 is composed of the rare words, which are often content-area related, ranging from 73,500 word families. Moreover, the band 9 under lexical resources criteria of IELTS Writing section indicates that the test takers should "use a wide range of vocabulary with very natural and sophisticated control of lexical features" (Higgins, 2017, p. 6). The Oxford International Business English Certificate (OIBEC), on the one hand, contains a vocabulary component, which sets the wide and advanced business and financial terminologies in students' written outputs (Douglas, 2000).

There are inconsistencies of vocabulary sophistication. Nation (1990) treated at least 2,000 to 3,000 words as high-frequency words. Stahl and Nagy (2005) believe that high-utility literate vocabulary, key content area vocabulary and high-frequency words deserve careful

attention for lexical sophistication. Achieving lexical sophistication may relate to what Stahl and Nagy (2005) commented that there are obstacles to large vocabulary acquisition, such as the case of the very large number of words in the English language; and that the academic English is not the same with the language used at home. Coxhead and Demecheleer (2018) also remind that “...conceptualization of what is technical in trades-based vocabulary is not straightforward, and that differences may be based on the experience of a trades-person, both in their trade and in their teaching” (p. 96). Lexical Frequency Profile (LFP) by Laufer and Nation (1995), which includes the most frequent 1000 words of English may have been subject to debate. Meara (2005) argues that LFP may be problematic. Two of the reasons include the concerns if error treatment in frequency profiling is counted or not; and whether formulaic and collocations such as ‘Victoria Park’ are counted as one or separate lexical item. Nonetheless, the General Service List of English Words (GSL) (West, 1953) can be used to indicate a level of lexical sophistication. What it means is that the low-frequency words are considered more sophisticated. These words are mostly academic and technical in nature (cf. Coxhead, 2000; Thawarom & Singhasiri, 2020). In fact, Ha (2019) also used GSL1, GSL2 and AWL in the text as a measure of lexical sophistication.

2.3.2. Lexical Density:

Read (2000) reviews that it was Ure (1971) who originated the measurement of lexical density. The calculation of lexical density starts with the identification of lexical items, which include nouns, verbs, adjectives, and adverbs of time, place and manner. Verbs *be* and *have*; auxiliaries, determiners, proforms, numerals, interrogative and negative adverbs, prepositions, conjunctions, discourse markers and interjectors are considered grammatical items (Crawford Camiciottoli, 2007). The identified number of lexical items is then divided by the total number of tokens in the text and is multiplied by 100 to yield a percentage of lexical density.

However, a definite average lexical density is far from settled. Stubbs (1996) maintained that written texts have a lexical density ranging from about 40-65%, which is comparable with Ure's (1971) calculation of more than 40%. For Samad (2004), the normal lexical density of written texts is a ratio ranging from 36-57%. Ure (1971) and Halliday (1985) also compared that written texts have higher lexical density than spoken texts, which is about less than 40%. However, Thawarom and Singhasiri (2020) computed that the lexical density of one-minute speaking task is 43.72%. They also found out the predictable dominance of nouns, given that they are ubiquitous in nature.

2.3.3. Lexical Diversity:

“Lexical diversity is a complex property that summarizes the range of vocabulary and the avoidance of repetition in the sample” (Malvern & Richards, 2013, p. 1). It is about the variety of different words used in the context (Read, 2000), and is usually employed to assess language in use (Malvern & Richards, 1997). The common way of measuring lexical diversity is the Type-Token Ratio (TTR), that is, the number of different words is divided by the total number of words. However, Malvern and Richards (1997) warn that the computation of TTR may be erroneous. Consequently, they listed different variations of TTR and other diversity measures, but still highlighted some of their disadvantages (cf. pp. 62-63). On the one hand, Malvern, Richards, Chipere, and Duran (2004) used Number of Different Words (NDW) as a measure of range, indicating how diverse the lexis can be.

There are concerns on diversity computation. “TTR is not a constant, but decreases with increasing size of token sample. The reason is straightforward as the number of tokens increases, the available pool of new types diminishes and the more tokens in any sample, the greater the probability of repetitions” (Malvern & Richards, 1997, p. 60). Read (2000) also reminds that proficient writers show the proclivity to use different words to avoid repetition. The variety of words can also be achieved using synonyms.

3. Methodology

3.1. Corpus

The data of this present study only reached 23,502 tokens. Higginbotham and Reid (2018) considered their corpus of 920,000 words of L2 essays as large-scale. Following their assertion, this study is rather small-scale. The use of small corpora has been upheld by leading scholars to date (cf. Barnbrook & Sinclair, 2001; Bondi, 2001; Ghadessy, Henry, & Roseberry, 2001; Ragan, 2001). Sinclair (2001) assures that a small corpus remains a relevant and reliable body of evidence. Anthony (2009) assures that the use of small corpus is also hugely consequential to the analysis of unique properties to glean from them— where he used the analogy of looking at the small group of stars. He continues that the detailed information, not the size of the corpus is what matters. Likewise, Baker (2006) maintains that 200,000 words may be the least number of words for a discourse analysis. This present study is not an attempt to look at the discursive patterns of the comments.

3.2. Corpus Producers and the Writing Task

The participants in this study were the 36 freshmen Filipino university students taking Tourism (13 females and 5 males) and Marketing (12 females and 6 males). They were officially enrolled in *Purposive Communication* course under me for the academic year 2019-2020-First Trimester, covering from 15 July 2019 to 12 October 2019. The Tourism cluster had an official total of 30 students. However, only 18 students commented on the complete 10 reading articles. To observe comparability, only 18 students from Marketing cluster who also completed 10 reading articles were randomly selected. Unfortunately, I could not secure the official proficiency levels of these students. The department does not have any official records of students' language proficiency levels, which are only left to discretion of the teachers alone (cf. Munalim, 2020, p. 289). Nonetheless, the comparative

undertaking in this present study may be defensible given that the students from different disciplines have different specific language needs and interests.

The analysis of students' lexical richness emanates from the reading-writing requirement of the course. Until now, all students are required to read posted online articles, and comment on them in one paragraph. This required number of paragraphs has been consistent since the inception in 2017. This reading-writing task serves as a preparation for a major writing project titled as "The Young Columnists," which requires the students to write at least 11 opinion articles (2 articles per week) for the remaining four weeks in the final term.

The social media worlds have been replete with too much information, and information-overload may precipitate unhealthy discourses. Brevity means respect to the readers. This is one of the standards I set for my *Purposive Communication (in English)* course, especially in their minor weekly task—reading of local and international articles, and then commenting on them in one paragraph. This writing restriction will help the students to filter only the most important information that matters in an ongoing media discourse. Inherent in these comments is the need for students to observe lexical richness under paragraph-limit pressure. Lexical richness is indicative of the students' informational contents, propositions and arguments in getting their messages across to the intended readers.

The writing and reading tasks are anchored on The Philippine Women's University's Inquiry-Experiential-Project-Problem-Transformative-Transformational Multicultural approach (cf. Munalim, 2019a; Munalim, 2019c). My classes engage around a specific matrix of problematique, which allows my students to question, cogitate, articulate personal ideas, and develop a disposition to question the world of things around them. The topics cover a range of personal, political, social, religious, and educational issues, to mention a few.

For the whole term, which runs for four months, students have to read and comment on 20 articles. For this present analysis, the first 10 articles of selected students were analyzed. Not all of them commented on all 20 articles. This strategic move was to achieve comparability. Malvern, Richards, Chipere, and Duran (2004) advice that making comparisons between groups, individuals, activities or contexts should demand identical sample sizes.

3.4. Data Collection and Analysis

Comments posted on our private Facebook groups posed easy retrieval. They were directly lifted from the groups without qualms of accuracy. The two sets of corpus were culled based on the original number of words/tokens. The corpus of Marketing cluster, which was longer than the Tourism cluster, was not truncated, thus may not affect the comparative study. Jarvis (2002) assures:

Because the D index represents the shape and position of the entire TTR curve, this index is assumed not to be affected by text length. That is, a higher curve—regardless of length—represents greater LDV than a lower curve does. The D index obtained through curve fitting should, thus, be ideal for comparing texts of different lengths. (p. 61) The following criteria were observed:

(1) Tagalog words were included, as they formed the local colors and Philippine Englishes of the students (cf. Munalim, 2019c). There were limited Tagalog words, given that the articles range from local and international scopes. These local words would not have had any appreciable and confounding effects on the expected results. This is assured, especially that all corpus tools used in this study, would analyze even the foreign words;

(2) Right spacing of the words, typos, misspelling, truncation, abbreviation, and other cases of cyber-orthography (cf. King, 2009) were re-checked so that the corpus tools to use could analyze them accurately. Proper nouns such as “Mayon Volcano” and Manila City Mayor “Isko Moreno” were changed to “Mayon_Volcano” and “Isko_Moreno” so that the corpus tools could read them as one lexical item; and

(3) Learner corpus was free from plagiarism, as also stipulated in the course house rules.

Corpus tools used include AntWordProfiler (Anthony, 2014) for measuring sophistication; TagAnt (Anthony, 2015) for culling lexical items; and D_Tools (Meara & Miralpeix, 2016) for measuring diversity. The computation of lexical density was coursed through the UAM Corpus Tool (O'Donnell, 2015) for instant POS-tagging and frequency counting. As assembled above, there are nuances of lexical sophistication, density and diversity measurement and profiling (cf. Hallidy, 1985; Jarvis, 2002; Laufer, 2005; Malvern, et al., 2004; Meara, 2005; Ure, 1971; Yu, 2009). The results in this present study were heavily based upon what these corpus tools can generate. Meanwhile, z test on two proportion samples was used with the aid of SPSS to see the significant differences of the variables under study. In comparing the two percentages, chi square test was utilized, as it is used to compare or test of associations if the data is at nominal level."

4. Results

This present study has been guided by two questions. In the interest of organization, the second question, which is an attempt to account for variations between two disciplines, has been embedded in the first question for the fluidity of the presentation.

4.1. Lexical Sophistication

4.1.1. *Vocabulary Profile:*

Vocabulary profile generated by AntWordProfiler (Anthony, 2014) consists of three built-in level lists such as General Service Lists (GSL_1st1000, GSL_2nd1000), and Academic Word List (AWL_570) including one non-based list. The non-based list represents the words that do not belong to the default lists, and is an optional output.

As divulged in Table 1, results foregrounded that both Tourism and Marketing students employ 19,865 tokens from the cohort of GSL1000 when writing article comments. There are 1,057 tokens under GLS2nd1000, while there are 923 tokens under the AWL. The percentage of token under GSL1st100 is 84.52%. Of the total tokens under GSL1000, there are 1,295 types of vocabulary; 419 types under GSL2000; and 345 vocabulary types under the AWL. Under the word families, 686 belonged to GSL1000. Meanwhile, there are 230 word families at the AWL level, which is almost closer to GSL2000 level of 282 word families.

Comparing the GSL1st100 percentage of 84.52% with the 80% coverage suggested by Nation (2001) for all high frequency words, there is reason to assert that the overall lexical sophistication of one-paragraph comments is lexically high. However, the percentage of the AWL under token only reaches 3.93%. This falls below Nation's (2001) proposed 10% coverage of the Academic Word List. This result indicates that there are few cases of academic lexical tokens in one-paragraph comments.

Table 1. Overall Lexical Sophistication in One-Paragraph Comments

Level	File	Token	%	Type	%	Group	%
1	1_gsl_1st_1000	19865	84.52	1295	48.14	686	37.51
2	2_gsl_2nd_1000	1057	4.50	419	15.58	282	15.42
3	3awl_570	923	3.93	345	12.83	230	12.58
0	-	1657	7.05	631	23.46	631	34.50
<i>Total</i>		23502		2690		1829	

The vocabulary profile is treated against the two clusters such as Tourism and Marketing. As presented in Table 2, the group of Marketing students perches on top in terms of token, type and word families. They employ more tokens than Tourism students. As “heavy senders” (cf. Claridge, 2007, p. 93), they are likely to employ more types and group families compared to Tourism students. Jarvis (2002) supports that “the more lexical tokens the learner uses, the more likely he or she is to repeat a lexical type” (p. 58). The same case has been noted in these one-paragraph comments.

Table 2. Comparison of Lexical Sophistication Between Clusters

Clusters	Level	File	Token	%	Type	%	Group	%
Tourism	1	1_gsl_1st1000	8999	84.43	890	54.74	540	46.08
	2	2_gsl_2nd1000	488	4.58	222	13.65	164	13.99
	3	3awl_570	370	3.47	187	11.50	141	12.03

	0	-	801	7.52	327	20.11	327	27.90
	<i>Total</i>		10658		1626		1172	
	1	1_gsl_1st1000	10866	84.60	1031	52.20	595	42.90
Marketing	2	2_gsl_2nd1000	569	4.43	311	15.75	228	16.44
	3	3awl_570	553	4.31	239	12.10	170	12.26
	0	-	856	6.66	394	19.95	394	28.41
	<i>Total</i>		12844		1975		1387	

In terms of token, both of the Tourism cluster and the Marketing cluster show no significant differences in the two levels of the 1st and 2nd 1000. Table 3 shows that the significant difference only lies at the third level of the AWL_570, where the Marketing cluster token percentage of 4.31% is higher. This means that students from the Marketing cluster employ more cases of academic words than the Tourism cluster. This is further vouched in the overall significant difference of -9.40 at 0.0001 p value.

Table 3. Comparison of Cluster's Token Across Levels

Levels	TM (%)	MK (%)	Chi Square	Diff	p value
1_gsl_1st1000	84.43	84.6	0.129	-0.17	0.7198
2_gsl_2nd1000	4.58	4.43	0.305	0.15	0.5807
3awl_570	3.47	4.31	10.888	-0.84	0.0010
-	7.52	6.66	6.574	0.86	0.0103
<i>Total</i>	45.3	54.7	205.87	-9.4	0.0001

In terms of type, Table 4 divulges that the differences of types across all levels are not statistically significant. The difference only becomes statistically significant when all levels are merged into one, where the Marketing cluster employs more tokens in their one-paragraph comments. The overall pattern is an indication that the students from the Marketing cluster exhibit the tendency to delimit the use of repetition of word types when writing one-paragraph comments in all three levels of the GSL2000 and the AWL.

Table 4. Comparison of Cluster's Type

Levels	TM (%)	MK (%)	Diff	Chi Square	p value
1_gsl_1st1000	54.74	52.2	2.54	2.311	0.1285
2_gsl_2nd1000	13.65	15.75	-2.1	3.118	0.0774
3awl_570	11.5	12.1	-0.6	0.308	0.5791
-	20.11	19.95	0.16	0.905	0.9050
<i>Total</i>	45.2	54.8	-9.6	33.558	0.0001

In like manner, the patterns of type are cascaded into the group family. There are no significant differences of the group in all three levels. The only significant difference occurs when all group or word families are classified as one. As shown in Table 5, the Marketing cluster employs more hits of group families than the Tourism cluster. This pattern echoes the pattern in Table 3 on word types, where Marketing students use fewer cases of type repetition.

Table 5. Comparison of Cluster's Group Family

Levels	TM (%)	MK (%)	Diff	Chi Square	p value
1_gsl_1st1000	46.08	42.9	3.18	2.602	0.1068
2_gsl_2nd1000	13.99	16.44	-2.45	2.938	0.0865
3awl_570	12.03	12.26	-0.23	0.8592	0.8592
-	27.9	28.41	-0.51	0.0816	0.7751
<i>Total</i>	45.8	54.2	-8.4	17.923	0.0001

When TTR ratio was computed, the Marketing cluster has a higher TTR than the Tourism cluster. This higher percentage of TTR (0.1538) is an indication of a higher rate of repetition of tokens compared to the Tourism students. This result is predictable given that they use more tokens. Table 6 shows this pattern.

Table 6. Types, Tokens, and Type-Token Ratios

Clusters	Type	Token	TTR
Tourism	1626	10658	0.1526
Marketing	1975	12844	0.1538

4.1.2. *Rare and Technical Words.*

Rare word diversity (RWD) is related to lexical sophistication. According to Malvern and Richards (2009), the calculation of rare words is based on the partition between rare and basic vocabulary. I followed these authors, who further claim that “rare word usage can be derived from information intrinsic to the data set under study” (p. 169). They continue that “... it is likely that on average there are more people who do not use the word than there are who do, we can consider the word to be rare; otherwise it would be basic vocabulary” (p. 169). I only culled the words with 1 hit/occurrence for each word. Results show that from both clusters, there are 1,334 rare words. The Tourism cluster has 849 rare words, and 1,040 rare words from the Marketing cluster.

Table 7. Most Commonly Used Words

Both Clusters (f)	Tourism (f)	Marketing (f)
technology (77)	manners (44)	technology (61)
accent (75)	accent (37)	accent (38)
etiquette (49)	etiquette (25)	Aetas (29)
culture (48)	technology (16)	etiquette (24)
Aetas (44)	Aetas (15)	state (23)

By contrast, Table 7 shows the five commonly used words from both clusters. However, by looking closely at these rare words and the top five commonly used words, these are not rare but rather ordinary words. Using my personal epistemic judgment and “domain knowledge” (cf. Chung & Nation, 2004, p. 253), these words are actually taken from the required reading articles. This pattern supports that “word learning will also be dependent on what thematic material the learner has been exposed to in textbooks and on word difficulty factors such as whether the words encountered are cognate or not” (Milton & Alexiou, 2009, p. 198). It may be argued that even if these sets of vocabulary are just taken from the articles, we can still claim that the students have attempted to incorporate these words as textual evidence in their comments. However, lexical sophistication is rather poor.

4.2. Lexical Density

Lexical density provides a measure of the proportion of lexical items in the text” (Johansson, 2008, p. 61). As computed by UAM Corpus Tool (O’Donnell, 2015), Table 8 highlights that the overall percentage of lexical density of 57.84%. Predictably, nouns dominate among the types of lexical items, given that they are ubiquitous in nature. Overall, the percentage of 57.84% is 3-point higher than the average of lexical density between 40-54%, as computed by Stubbs (1996), and is just close to the computation of Samad (2004), who mentions that 36-57% is considered a normal lexical density. This pattern provides evidence that the students in their desire to lump ideas into one paragraph still manage to produce comments which are lexically dense.

Table 8. Overall Percentages of Lexical Density of Two Clusters

Domains	Total	N	V	ADJ	ADV	Total Content Words
	tokens					
No. of tokens	23502	5052	4886	1905	1751	13594
% of LD		21.50%	20.79%	8.11%	7.45%	57.84%

When an inferential statistics was employed, results show that there exists a significant difference between the four lexical items. Table 9 reveals that nouns turn out to have a significantly higher percentage of 21.5% as compared to adjectives (8.11%) and adverbs (7.45%). Percentage of nouns turns out to be not significantly different when compared to the resulting percentage of verbs (20.79%). All other pairwise comparisons turn out to be significant, where the verb percentage of 20.79% is higher than the adjective and adverb. Lastly, resulting percentage of 8.11% for the adjective is also significantly higher than 7.45% adverb.

Table 9. Significant Difference Among the Four Lexical Items

Items	%	Items	%	Difference	Chi Square	p value
N	21.5	V	20.79	0.71	3.553	0.1517
		ADJ	8.11	13.39	1670.337	0.0001
		ADV	7.45	14.05	1873.728	0.0001
V	20.79	ADJ	8.11	12.68	1528.327	0.0001
		ADV	7.45	13.34	1724.45	0.0001
ADJ	8.11	ADV	7.45	0.66	7.134	0.0422

Table 10 reveals the comparison of lexical density between the two clusters. It shows that the Marketing cluster still employs more hits of lexical items (58.25%) than the Tourism cluster (57.36%). These percentages are both considered at the peak of lexical density between 36-57% (Samad, 2004).

Table 10. Comparison of Lexical Density

Clusters	Domains	Total tokens	N	V	ADJ	ADV	Total Content Words
Tourism	No. of tokens	10,658	2237	2250	832	794	6113
	% of LD		20.99%	21.11%	7.81%	7.45%	57.36
Marketing	No. of tokens	12,844	2815	2636	1073	957	7481
	% of LD		21.92%	20.52%	8.35	7.45%	58.25

However, Table 11 shows that the resulting p value, which all exceeds 0.05, denotes that the Tourism and the Marketing clusters have the same percentage in terms of noun, verbs, adjective and adverbs. Statistically, both clusters tend to use lexical items identically even if the Marketing students tend to use more lexical items. Hence, the overall lexical density of one-paragraph comments is high (57.84%), as also reported in Table 8. By and large, the comments may have high proposition content and great complexity because of the high level of lexical density (cf. Crawford Camiciottoli, 2007).

Table 11. Significant Difference of Lexical Items in Both Clusters

Items	Tourism (%)	Marketing (%)	Diff.	Chi Square	p value
N	20.99	21.92	-0.93	2.985	0.1892
V	21.11	20.52	0.59	1.231	0.3991
ADJ	7.81	8.35	-0.54	2.280	0.2512
ADV	7.45	7.45	0.00	0.000	1.000

4.3. Lexical Diversity

4.3.1. Overall Lexical Diversity:

Unfortunately, D_Tools failed to process the present corpus of 23,502 tokens after feeding it within two weeks in three different geographical locations– two in the Philippines and one in Ohio, USA. Therefore, the overall lexical diversity cannot be presented here. In lieu of the overall group’s computation of lexical diversity, each cluster’s result is offered accordingly. Nonetheless, a cluster individual report can offer insights in terms of “how many different words are used in a text” (Johansson, 2008, p. 61).

**Figure 1. Failure of D_Tools to Process the Corpus of 23,502 Tokens**

4.3.2. Lexical Diversity of Two Clusters:

Figures 2 and 3 generated by D_Tools show the comparison of the mean segmental TTR values (the blue line) for samples of 35 to 50 words and the values generated by the best-fitting value of D (the yellow line) based upon the theoretical model by Malvern and Richards. Accordingly, the value of D can vary between 1 and 120. Low values of D indicate that the text under analysis contains more cases of repetition, thus lexically poor. On the one hand, high values of D are an indication that the text under investigation is lexically rich, with fewer cases of repetition of words. An error statistics shown in Figure 2 and 3 indicates how close the data matches the model. This figure of error should not be bigger than 0.01 (cf. Meara & Miralpeix, 2016).

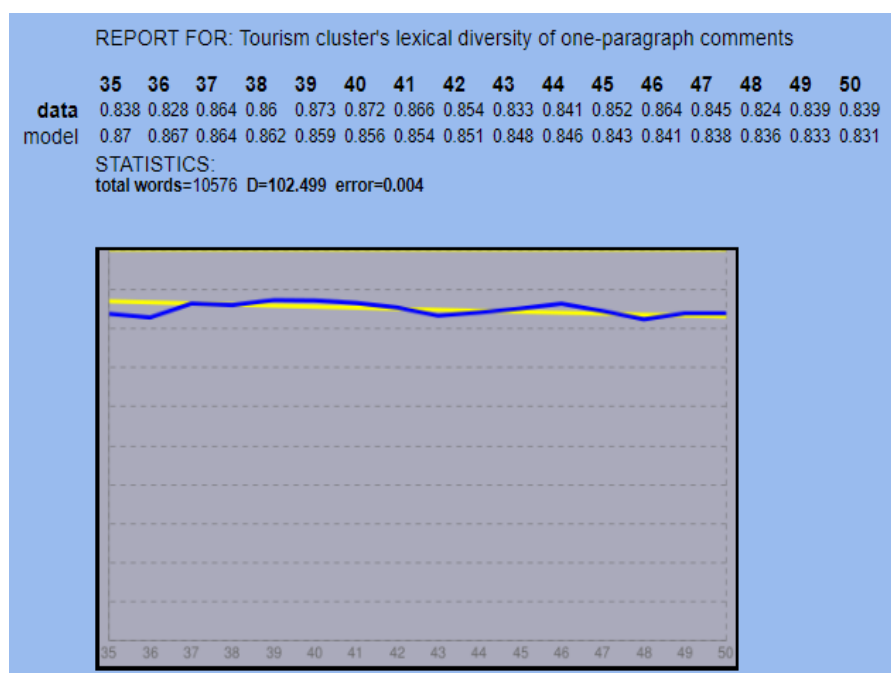


Figure 2: Tourism Cluster's Lexical Diversity of One-Paragraph Comments

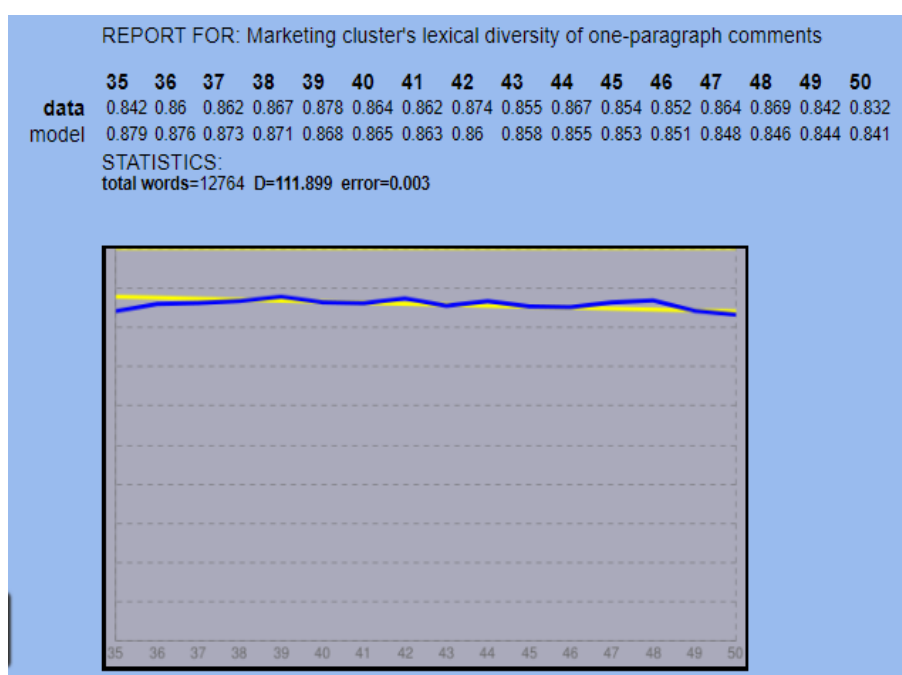


Figure 3. Marketing Cluster's Lexical Diversity of One-Paragraph Comments

Upon inspection of Figures 2 and 3, the value of D (yellow line) of the Tourism cluster is 102.499, while the value of D of the Marketing cluster is 111.899. The Marketing cluster then has richer lexical diversity. Jarvis (2002) explains that “a higher number represents greater LDV, and a lower number represents less LDV” (p. 61). To reiterate, the model lexical diversity is between 1 to 120. The average D value of two clusters is 107.199. As a group, the students' lexical diversity is about 89.16%, which may be considered high.

At a comparative level, Marketing students show the propensity to be more lexically diverse and richer (93.25%) than the Tourism students (85.42%). The Marketing cluster almost attaches closer to the highest D of 120 – with only a difference of 8.101. The pattern might be predictable because longer discourses are expected to generate greater number of different words (Malvern & Richards, 1997). The overall tokens of the Marketing group reaches 12,844 compared to 10,658 of the Tourism cluster. Jarvis (2002) however, clarifies that “this index is assumed not to be affected by text length. That is, a higher curve—regardless of length—represents greater LDV than a lower curve does. The D index

obtained through curve fitting should, thus, be ideal for comparing texts of different lengths” (p. 61).

Meanwhile, the error statistics of the Tourism cluster is 0.004, while Marketing cluster’s error statistics is 0.003– both are lower than 0.01. These figures of error statistics convey that the actual data from two clusters match the model data, and that the program is appropriate for current data analysis.

5. Overall Discussion

This present exploratory study offers the following patterns when Tourism and Marketing students wrote one-paragraph comments on our Facebook groups. On lexical sophistication, the whole one-paragraph comments are lexically sophisticated mostly at the GSL_1st100, which breaches more than the 80% highest mark suggested by Nation (2001). However, the comments are lexically poor in terms of academic words. This pattern is supported by lexical poorness in terms of the technicality and rareness of words. At the comparative level, the Marketing cluster is statistically consistent with more cases of token, type and word families. On lexical density, the whole one-paragraph comments are lexically dense (57.84%). Nouns are statistically higher than verbs, adjectives and adverbs. Although the Marketing cluster tends to use more lexical items, the difference turns out to be statistically not significant. Simply put, the results indicate that the students’ comments have high proposition content and great complexity even in just one paragraph. In terms of lexical diversity, the whole one-paragraph comments are lexically diverse (89.16%), which reaches the D value of 107.199 out of 120 highest D value. At the comparative level, the value of D of the Tourism cluster is 102.499, while the value of D of the Marketing cluster is 111.899. Students from the Marketing cluster shows some proclivity to be more lexically diverse (93.25%) than the students from the Tourism cluster (85.42%).

The overall lexical richness in one-paragraph comments are high, except for the poorness in academic and technical words. What this means is that students' writing may not be encouraged to write extended paragraphs if lexical richness is achieved. However, at the same time, this paper fails to determine the quality of their written comments, and should be left for future studies. Likewise, the comparative analysis in this study is moored in the remit of ESP, with discipline-based tuition-learning processes. Results show that lexical richness is discipline-induced. The Marketing students perform better than the Tourism students. At least from my epistemic stance as their course professor, the results may be predictable. Tourism students had shown the least performance both in other writing and speaking tasks, as compared to the other sections during this term. It is argued that regardless of disciplines, lexical sophistication, density and diversity should be rich. Knowing the differences of lexical richness is relevant to the ESP context because these differences inform the ESP teachers about how the students attempt to express meaning and employ written strategies in various communicative events (cf. Vedder & Benigno, 2016), in this study, writing one-paragraph comments on Facebook, which is a social platform for their ideas. Consequently, the students' lexical richness is intuitively useful for the four-fold pillars of an ESP classroom such as evaluation; materials and methods; needs analysis; and learning objectives in general, and students' own necessities, lacks, and wants for an ESP course in particular (cf. Anthony, 2018).

6. Conclusion

Students are expected to acquire a set of writing skills for social media discourses with brevity, but with high level of lexical sophistication, density and diversity. Within the students' ability to observe the vocabulary richness usually employs a range of skills to trim down ideas succinctly. The restricted one-paragraph comments on Facebook, as a type of register (cf. Crawford Camiciottoli, 2007; Crawford & Csomay, 2016), can of course affect the lexical richness, but this is not reason or excuse for students' lexical poorness. Likewise, they commented on the same articles, but this should not be the reason for lexical poorness because the English vocabulary is rich. It is in these results that support and shape

the ways in which teachers tend to view how student-writers manage to express in writing amid restrictions of the number of paragraphs, and how to trim kernel ideas succinctly, therefore the need for learners' need analysis in an ESP classroom (cf. Flowerdew, 2013).

It was not the mission of my paper to compare between one and more paragraphs. Future researchers may work on this before we can prudently argue whether longer written Facebook comments are indicative of lexical richness. Students' academic performance can be also a fertile future study within in the context of word-number restricted written outputs. The premise of these restricted one-paragraph comments is also to discourage the students to be roundabout. It is therefore a good research desideratum to look at the degree of digression in their comments. Likewise, we might as well ascertain whether the use of vocabulary is hastened by their personal perceptions of the need to be just simple, without using highfalutin words. This is far from settled in this present study. Moreover, the current study fails to address a number of variables related to the students and their proficiency backgrounds. For instance, future researchers may wish to consider the relationship between lexical richness and the quality of the writers' word-number-restricted comments.

Underpinning all of these results are a number of implications for ESP teachers' roles in need analysis, syllabus design, materials writing, adaptation, and evaluation in vocabulary and writing instruction (cf. Guest, 2016; Hutchinson & Waters, 1991; Nation, 2008; Stahl & Kapinus, 2001; Thompson, 2018). For instance, teachers need not require students with longer and extended paragraphs if the richness of vocabulary is still achieved at a higher level. Instead of quantity, teachers need to focus more on the quality of paragraphs. In due time, teachers can add the number of paragraphs incrementally once students' reading and writing stamina have improved (cf. Munalim, 2019b). Teachers and students can engage in one-on-one sessions to identify students' struggles, challenges, fluency and accuracy of the texts. They can also engage students in 'vocabulary rehab' (cf. Sprenger, 2014) because the amount of instruction can significantly increase lexical richness (Jarvis, 2002). To this end, teachers are encouraged to further their efforts in improving students writing skills with their lofty goals of preparing them to be ready in the outside and the real world of social

media discourses and technologized communication (cf. Flowerdew, 2013), including professional communication under the specific industries that they would wish to be inducted.

Acknowledgment and Declaration

I would like to thank the expert reviewers for their compelling suggestions; to the Associate Editor Prof. El-Sakran and Chief Editor Prof. Nunn for the patience and consideration for the delayed revision process, which was precipitated by the pandemic; the developers of helpful corpus tools used in this study; and Mr. Reginald P. Arimado for the statistical computation. Paper was done without support from any funding agency.

References

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425-438.
- Alderson, J.C. (2005). *Diagnosing foreign language proficiency*. London: Continuum.
- Anthony, L. (2018). *Introducing English for Specific Purposes*. London: Routledge.
- Anthony, L. (2015). *TagAnt (Version 1.2.0) [Computer Software]*. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software>
- Anthony, L. (2014). *AntWordProfiler (Version 1.4.0) [Computer Software]*. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software>
- Anthony, L. (2009). Issues in the design and development of software tools for corpus studies: the case for collaboration. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 87-104). London: Continuum.
- Baker, P. (2006). *Using corpora in discourse analysis*. New York: Continuum.
- Barnbrook, G., & Sinclair, J. (2001). Specialised corpus, local and functional grammars. In M.A. Ghadessy & R.L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 237-276). Amsterdam: John Benjamins.
- Beck, I.L., McKeown, M.G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.

- Beck, I.L., McKeown, M.G., & Kucan, L. (2005). Choosing words to teach. In E.H. Hiebert & M.L. Kamil (Eds.), *Teaching and learning vocabulary* (pp. 209-222). New Jersey: Lawrence Erlbaum.
- Benesch, S. (2001). *Critical English for academic purposes: Theory, politics, and practice*. New Jersey: Lawrence Erlbaum.
- Bloch, J. (2013). Technology and ESP. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 387-401). West Sussex: John Wiley & Sons.
- Block, D. (2019). *Post-truth and political discourse*. London: Palgrave Macmillan.
- Brown, J.D. (2016). *Introducing needs analysis and English for Specific Purposes*. London, UK: Routledge.
- Cambridge University Press (CUP). (2005). *Cambridge IELTS 4: Examination papers from University of Cambridge ESOL examinations: English for speakers of other languages*. Cambridge: Cambridge University Press.
- Carter, R. (2012). *Vocabulary: Applied linguistics perspectives*. London: Routledge.
- Christakis, N.A., & Fowler, J. H. (2011). *Connected: The surprising power of our social networks and how they shape our lives—How your friends' friends' friends affect everything you feel, think, and do*. NY: Back Bay Books/Little, Brown and Company.
- Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System*, 32(2), 251-263.

- Claridge, C. (2007). Constructing a corpus from the web: Message boards. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 87-108). Amsterdam: Rodopi.
- Clark, M.C., Hergenrader, T., & Rein, J. (Eds.). (2015). *Creative writing in the digital age: Theory, practice, and pedagogy*. USA: Bloomsbury Publishing.
- Clyne, M. (2002) Contrastive discourse studies. *Cahiers de praxématique* 38. Montpellier: Pulm. 59-84.
- Cooper, P. A. (1993). An analysis of lexical richness in student writing. *Language Matters*, 24(1), 147-163.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Coxhead, A. (2013). Vocabulary and ESP. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 115-132). West Sussex: John Wiley & Sons.
- Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of Plumbing. *English for Specific Purposes*, 51, 84-97.
- Crawford Camiciottoli, B. (2007). *The language of business studies lectures: A corpus-assisted analysis*. Amsterdam: John Benjamins.
- Crawford, W.J., & Csomay, E. (2016). *Doing corpus linguistics*. London: Routledge.

- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197-222.
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66-76.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge, UK: Cambridge University Press.
- Dudley-Evans, T., & St. John, M.J. (1998). *Development in English for specific purposes: A multi-disciplinary approach*. UK: Cambridge University Press.
- Enriquez, V.G. (1992). *From colonial to liberation psychology*. Quezon City: University of the Philippines Press.
- Flowerdew, L. (2013). Needs analysis and curriculum development in ESP. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 325-346). West Sussex: John Wiley & Sons.
- Ghadessy, M., Henry, A., & Roseberry, R.L. (2001). *Small corpus studies and ELT: Theory and practice*. Amsterdam: John Benjamins.
- Guest, M. (2016). Overcoming institutional barriers to establishing an ESP programme: A case report from Japan. *The Asian ESP Journal*, 12(3), 7-24.
- Halliday, M.A.K. (1985). *Spoken and written language*. Oxford: Oxford University Press.

- Higginbotham, G., & Reid, J. (2018). The lexical sophistication of second language learner's academic essay. *Journal of English for Academic Purposes*, 37, 127-140.
- Higgins, R.T. (2017). *IELTS: Academic and general task 2: How to write at a band 9 level*. IELTS Materials/ Unique Materials.
- Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes*, 33, 54-65.
- Hutchinson, T., & Waters, A. (1991). *English for specific purposes: A learning-centered approach*. Cambridge: Cambridge University Press.
- Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57-84.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working Papers*, 53, 61-79.
- Johns, A.M. (1997). *Text, role and context. Developing academic literacies*. Cambridge: Cambridge University Press.
- Kaplan, R.B. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16, 1-20.
- Kellner, D. (2000). New technologies/new literacies: Reconstructing education for the new millennium. *Philosophy of Education*. Retrieved from <https://pages.gseis.ucla.edu/faculty/.../newtechnologiesnewliteracies.pdf>

- King, B. (2009). Building and analysing corpora of computer-mediated communication. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 301-320). London: Continuum.
- Krashen, S.D. (1981). *Principle and practice in second language acquisition*. UK: Pergamon Press.
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757-786.
- Laufer, B. (2005). Lexical frequency profiles: From Monte Carlo to the Real World: A Response to Mearca. *Applied Linguistics*, 26(1), 32-47.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16, 307-22.
- Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71) Clevedon: Multilingual Matters.
- Malvern, D. D., & Richards, B., J. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85-104.

- Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills, NH: Palgrave Macmillan.
- Malvern, D., & Richards, B. (2009). A new method of measuring rare word diversity: The example of L2 learners of French. In B. Richards, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 164-178). Hampshire: Palgrave Macmillan.
- Malvern, D., & Richards, B. (2013). Measures of lexical richness. In C.A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-5). London: Blackwell.
- Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development. Quantification and assessment*. Basingstoke, Hampshire: Palgrave Macmillan.
- Malvern, D.D., & Richards, B.J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71) Clevedon: Multilingual Matters.
- Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics*, 26(1), 32-47.
- Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary*. Clevedon: Multilingual Matters.
- Milton, J., & Alexiou, T. (2009). Vocabulary size and the Common European Framework of References for Languages. In B. Richards, D. D. Malvern, P. Meara, J. Milton, & J.

- Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 194-211). Hampshire: Palgrave MacMillan.
- Milton, J., & Hales, T. (1997). Applying a lexical profiling system to technical English. In A. Ryan, & A. Wray (Eds.), *Evolving models of language* (pp. 72-83). Clevedon: Multilingual Matters.
- Munalim, L.O. (2020). Students as test makers in Spanish as foreign language: Insights for teacher-student collaboration in test constructions. *Journal of Foreign Language Teaching*, 17(1), 287-300.
- Munalim, L.O. (2019a). Micro and macro practices of multicultural education in a Philippine university: Is it global integration ready? *The Asia-Pacific Education Researcher*, 29, 441-454. <https://link.springer.com/article/10.1007/s40299-019-00497-7>
- Munalim, L.O. (2019b). Micro-level text contents of one-to-five minute news: American and Philippine English compared. *Journal of English as an International Language*, 14(2), 44-71.
- Munalim, L.O. (2019c). Subject-auxiliary inversion in embedded questions in spoken professional discourses: A comparison of Philippine English between 1999 and 2016-2019. *Journal of English as an International Language*, 14(1), 40-57.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. Florence, KY: Heinle & Heinle.

- Nation, P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston: Heinle Cengage.
- O'Donnell, M. (2015). *UAM corpus tool (Version 3.2j)* [Computer Software]. Available on <http://www.wagsoft.com/CorpusTool>
- Paltridge, B. (2013). Genre and English for Specific Purposes. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 347-366). West Sussex: John Wiley & Sons.
- Paltridge, B. & Starfield, S. (Eds.). (2012). *The handbook of English for specific purposes* (Volume 120). London: John Wiley & Sons.
- Pe-Pua, R., & Protacio-Marcelino, E. (2000). Sikolohiyang Pilipino (Filipino psychology): A legacy of Virgilio G. Enriquez. *Asian Journal of Social Psychology*, 3, 49-71.
- Ragan, P.H. (2001). Classroom use of a systemic functional small learner corpus. In M.A. Ghadessy & R.L. Roseberry, (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 207-236). Amsterdam: John Benjamins.
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.
- Samad, A.A. (2004). Beyond concordance lines: Using concordances to investigate language development. *Internet Journal of e Language Learning & Teaching*, 1(1), 43-51.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave Macmillan.

- Sinclair, J. Mc. 2001. Preface. In M.A. Ghadessy & A. Henry (Eds), *Small corpus studies and ELT: Theory and practice* (pp. vii-xv). Amsterdam/Philadelphia: John Benjamins.
- Smith, J. A., & Kelly, C. (2002). Stylistic constancy and change across literary corpora: using measures of lexical richness to date works. *Computers and the Humanities*, 36, 411–30.
- Sprenger, M. (2014). *Vocab rehab: How do I teach vocabulary effectively with limited time?* Alexandria, VA: ASCD.
- Stahl, S.A., & Kapinus, B.A. (2001). *Word power: What every educator needs to know about teaching vocabulary*. Washington, DC: NEA.
- Starfield, S. (2013). Critical perspectives on ESP. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 461-479). West Sussex: John Wiley & Sons.
- Stubbs, M. 1996. *Text and corpus analysis. Computer-assisted studies of language and culture*. London: Blackwell.
- Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. London: Routledge.
- Thawarom, T., & Singhasiri, W. (2020). Lexical richness of one-minute speaking task by science and technology university students. *The Journal of Asia TEFL*, 17(1), 70-86.

- Thompson, G. (2018). Insights for efficacy development from an exploration of Japanese business management students' EAP self-efficacy beliefs. *The Asian ESP Journal*, 14(7.1), 244 – 284.
- Todd, R.W. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, 45, 31-39.
- Ure, J. (1971). Lexical density and register differentiation. In G.E. Perren & J.L.M. Trim (Eds.), *Applications in Linguistics* (pp. 43–452). London: Cambridge University Press.
- Vedder, I., & Benigno, V. (2016). Lexical richness and collocational competence in second-language writing. *IRAL*, 54(1), 23-42.
- West, M. (1953). *A general service list of English words*. London: Longman.
- Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied linguistics*, 31(2), 236-259.