

Vocabulary Profile of the Abstracts Written by Students Majoring in Linguistics, Literature, and Cultural Studies: A Comparative Analysis

Imamatul Khair
Universitas Airlangga
Surabaya, Indonesia
imakhair@gmail.com

Ahmad Faris Firdaws S
Universitas Airlangga
Surabaya, Indonesia

Risky Amalia Ramdhani
Universitas Airlangga
Surabaya, Indonesia
risky.amalia.ramdhani-2014@fib.unair.ac.id

Sindy Alicia Gunawan
Universitas Airlangga
Surabaya, Indonesia
sindy.alicia.gunawan-2018@fib.unair.ac.id

Abstract—Vocabulary profile can be used to identify the uniqueness of certain texts among sub-genres. In this study, some types of vocabularies are examined. The study aimed at identifying vocabulary profile consisting of General Service List (GSL) and Academic Word List (AWL) in three different sub-genres of English major. This study used a qualitative method in identifying the elaboration on frequency of GSL and AWL as well as the dominant words emerging in 30 thesis abstracts. The thesis abstracts were retrieved from Linguistics, Cultural Studies, and Literature sub-genres. Only abstracts with A or AB score published from 2015 to 2016 were used. Based on types and tokens of GSL, it was found that Linguistics and Literature have higher number than Cultural Studies. For overall GSL, Linguistics covers more GSL compared to the other two sub-genres. One of reasons to this case is because there are many plural forms of the same words used in the texts. In terms of AWL, Cultural Studies possesses the highest number of AWL. It is because of the objects and scope of the sub-genres. In conclusion, among three sub-genres studied, Linguistics has the highest number of GSL meanwhile Cultural Studies are dominant for the AWL. The study gives implications to the writing skill for a specific text by considering types of words used.

Keywords—General Service List; Academic Word List; technical words; low-frequency words; types; tokens

I. INTRODUCTION

Vocabulary profile can infer information. Deriving from this case, many scholars have taken the advantage of making use of vocabulary profile. It can be used for evaluating the vocabulary growth especially in children as shown in Plunkett's work [1]. It is also useful to be used in measuring academic performance as shown in Morris and Cobb's study [2]. Last but not least, it can also be used to distinguish one type of text to another. A simple example to help us relate to this function is that an academic text, if compared to text in a

magazine or online news, surely would show different vocabulary profile.

Academic text itself has a lot of sub-genres in which each sub-genre has their own dynamic. The uniqueness of each sub-genre relies on the lexical choices, especially the vocabulary profile within the text. For example, the academic vocabulary in chemistry research articles used as the object of their study to have 9.60% coverage for the total corpus data [3]. Meanwhile, the academic vocabulary in agriculture research articles used the data for their study to cover 9.06% of the text [4].

Vocabulary is distinguished into four types: high-frequency words, academic words, technical words, and low-frequency words [5]. High-frequency words refer to the general words frequently used in the text. Technical words refer to the words which belong to a particular topic or subject area in which the words are only common in the texts covering the topic or subject area. The last one is low-frequency words which refer to the words that are excluded from the previous types. As suggested by its name, the words in this vocabulary type are rarely used in texts.

This study aims to examine the vocabulary profile within the abstracts of undergraduate thesis by students from Linguistics, Cultural Studies, and Literature major in English Department of a particular university. This study is conducted basically to contribute in enriching corpus linguistic studies especially about vocabulary profile in academic texts. We really hope by showing a report of the academic works produced by undergraduate students of English Department of a particular university, students from indifferent major could have a reflection to help them improve the quality of the lexical choice on their works based on the finding. To add, this study could also serve as the basis for future works especially which discuss the same topic in social science scope.

II. METHODS

The writers of this study apply qualitative approach that aims at giving broader interpretation [6]. The quantitative data in this study are interpreted and generalized by using qualitative approach. Even though most of corpora are analyzed by using quantitative approach, the qualitative analysis is not trivial. The writers decided to compare the Academic World List and General Service List from English Department students' thesis abstracts of a particular university in Surabaya. Abstract is selected to be the genre of text because 1) Shorter texts alleviate bias which commonly exists in longer texts because of the influence of word accounts. 2) Short texts cover more various kinds of topics, while longer texts often discuss about one topic. 3) The source of data is collected from several different authors. It contributes to give more balanced and unbiased vocabulary list used in the abstracts.

There are three steps to collect data that are required in this study. First, the writers looked for list of theses' titles that can be accessed on <http://repository.unair.ac.id>. Then, the writers sorted the titles by clicking Browse Repository bar which shows several tabs. After that, the writers browsed the data on Ilmu Budaya tab and chose Sastra Inggris column. The website showed several lists of publication years, but the writers only sorted the titles from year 2015 and 2016. Second, to know the scores of the theses, the writer checked the hardcopy of the theses in Thesis Room provided by the university library. Third, after making some notes for the selected titles, the writers accessed Repository website again and downloaded the abstract which can be accessed online.

For corpus analysis, the writers used a vocabulary profiler called AntWord. The analysis for word family was espoused. For example, the word families of "accept" have different word forms, such as *acceptance*, *acceptable*, *accept(s)*, *accepted*. The AntWord software is used to compare the vocabulary profile of 30 abstracts within Cultural Studies, Linguistics, and Literature. It has three different levels. Level 1 and 2 include 1000 most frequently general words in GSL. Level 3 covers 570 most frequent words in AWL, and level 4 contains unfound words in the AWL and GSL Database.

III. RESULTS AND DISCUSSIONS

A. Type Distribution

The words are representing the whole idea of academic work. They also represent the knowledge of the writer. Every student is different one to another in their ability to construct words within a single sentence. In one sense of 'word' we may count three different words; in another sense we may count ten different words. Words in the first sense are called "types" and words in the second sense "tokens". The type and token distribution in high frequency and low frequency words may different between linguistic, Cultural Studies, and literature.

TABLE I. TYPE OF TOKEN DISTRIBUTION IN LINGUISTICS, CULTURAL STUDIES, AND LITERATURE

Level	Linguistic		Cultural Studies		Literature	
	Token %	Type %	Token %	Type %	Token %	Type %
Level 1	72.77	34.21	66.89	45.99	71.3	32.97
Level 2	6.39	9.84	5.33	9.27	3.47	6.8
Level 3	9.17	13.26	12.4	17.17	9.15	15.92
Level 0	11.67	20.68	14.31	27.57	16.08	24.31

Table 1 shows the percentage of token and type words. The result differs between the entire majors. If we compare the token and type of the first and second level of words, the largest ratio between token and type is coming from Literature with 1.88 and followed by Linguistic with the ratio of 1.79 while Cultural Studies has the least number of ratio with 1.2. The level 3 shows that the use of type words is more often than token but the similar result appears in level 3, Literature has the largest ratio between token and type with 1.73 while Linguistic is in the number of 1.44 and followed by Cultural Studies with 0.13. According to the data, the used of type words in Literature are more frequent than any other major. The Linguistic are followed with almost similar number but Cultural Studies as the least ratio has the quite far number compare to other majors. It can be inferred that both Linguistic and Literature students mostly have the great knowledge on the choice of word. The students of Literature and Linguistic tend to be more creative than the students of Cultural Studies.

Different result is shown in the Low Frequency word list. The ratio between the token and type in the linguistic are larger than others with the number of 1.77. Followed by Cultural Studies with 1.9 and the least number is literature with 1.51. These results are contradicted with the high frequency level. It may refer to the fact that many of literature abstract are not talking about specific term of area. They have the least of low frequency level (which is going to be explained further on the other point). So, the use of token word in Literature will be less than other majors.

B. General Service List in Linguistics, Cultural Studies, and Literature

In this section, the GSL result for English theses' abstracts is presented. The result covers words that are mostly used in each of sub-genres by including tokens or running words, types or different words, word family as well as the interpretation of the profiles. GSL can be considered as the high frequency of words.

From the preview of each vocabulary profile, it is found that the accumulated GSL percentage in the first and second level of each sub-genre has different values. The highest number of the most frequent words belongs to Linguistics. It has 72.80% GSL in the first level and 6.40% in the second level. Literature abstracts have 71.30% GSL coverage in the first level and 3.5% in the second level. Compared to

Linguistics and Literature, Cultural Studies have covered 66.90% in the first level and 5.30% in the second level, meaning that the coverage involves the least number of general words.

TABLE II. PERCENTAGE DISTRIBUTION OF GENERAL SERVICE LIST

Level	Linguistics	Cultural Studies	Literature
Level 1	72.80%	66.90%	71.30%
Level 2	6.40%	5.30%	3.50%
Level 3	9.20%	12.40%	9.10%
Level 4	11.70%	15.40%	16.10%

Comparing these three sub-genres in terms of GSL, it raises interesting result which shows Linguistics instead have the highest number of high frequency words. It has 79.2 % of high frequency words, while Cultural Studies only have 72.2% and Literature 74.8%. This finding shows us that probably there are some differences in terms of word choice, word count, or topic that the sub-genres have. By identifying the tokens and types of words within the sub-genres, it was found that Linguistics has the highest number of tokens, 1915 tokens in total. The tokens and types in each sub-genre are counted from level 1 and level 2. In terms of the types, Linguistics has 447 types of words. The types of words in Linguistics only involve 23.3% of the tokens. However, Cultural Studies and Literature do not really differ in terms of the type word percentage of the tokens, which takes 26.2% and 26.1% respectively. Having this comparison of the tokens and types, it can be inferred that Linguistics has the smallest word types in GSL, only 23.3%, but the running words of GSL here are found to be much higher than Literature and Cultural Studies. With 23% types out of 1915 tokens, Linguistics does not really use various words of GSL. It still has the highest use of GSL even though the word types are less. This is because there probably are a number of words that are used repeatedly.

TABLE III. FREQUENCY DISTRIBUTION OF TOKEN AND TYPE IN THE GSL LIST

Linguistics		Literature		Cultural Studies	
the	217	the	188	the	170
of	98	of	79	of	107
in	86	in	72	in	78
and	72	and	59	and	71
to	64	to	59	to	55
is	55	is	48	is	45
that	33	this	39	as	41
this	27	a	28	a	31
are	26	study	28	this	31
by	26	that	28	that	25

This reason is further explained by looking at the 10 top list of words found in the vocabulary profiler. Linguistics covers the following use of words. Based on the table below, the words in level 1 GSL that are mostly used by all of sub-genres are prepositions, to-be, and determiner. The number of preposition, to-be, and determiner is higher than other words because these words follow other words, such as noun and verb. However, the use of those words in each sub-genre is different in terms of number. For example, determiner “the” in Linguistics appears 217 times, while Literature and Cultural

Studies use that for 188 and 170 times respectively. It is the same as the use of the preposition “in” in Linguistics, Literature, and Cultural Studies. The highest repetitive use of the word “in” belongs to Linguistics. Even though the number of some words of GSL in Linguistics, Literature, and Cultural Studies does not differ too much, the biggest contribution of GSL, in this case, comes from the use of determiner “the”. The gain difference of Linguistics and Literature, and Linguistics and Cultural Studies is respectively 29 and 47 “the” repetition. Because Linguistics tends to use high frequency words repetitively, it also influences overall percentage of GSL.

Furthermore, level 2 shows more GSL words besides general words in level 1. In this level, the high frequency words refer to be more specific and higher level, but they are still categorized as GSL. The table below provides the 10 top list of words in level 2 within each sub-genre. Linguistics, however, covers a lot more use of words in level 2 categories compared to Cultural Studies and Literature. For example, the word “noun” in Linguistics is repeated for 14 times. If it is compared to Literature, the repetition of word “aims” found in the first list of level 2 categories appears only six times. What makes Linguistics has more GSL coverage is due to the use of plural form of the same word. For example, the word “female” in Linguistics appears for 7 times. Then, there are the words “females” which are also counted in different section. The word “female” and “females” are separately counted. Thus, the variation of words will give more number of GSL usages. In short, the more words are repeated, the higher percentage of GSL will be.

Other than that, Linguistics provides more the use of general words compared to two other sub-genres. The core object in Linguistics is the language itself. It does not discuss the pervasiveness of cultural phenomena, but they emphasize more on one topic which is the language itself. In a sense of this aspect, Linguistics often studies about the abstract elements in language rather than in the culture. For example, level 1 also consists of some words that are not dominant compared to the 10 top list that has been stated earlier. Some of the words involve “learning,” “language,” and “English.” These words are closely related to topics that Linguistics commonly studies. The repetition of the words is more than fifteen times.

Linguistics also discusses about the morphology, syntax, or phonology of English language, so it does not reach the bigger scope in which many theories should be employed. The topic discussion is static, unlike cultural studies. The simple example can be taken from the analysis of signs in malls. The signs are not new things in our everyday life, but Linguistics has more roles in interpreting the way the signs contribute to the cultural construction among society. Thus, it tends to use words that exist for daily use like in level 2 categories. For example, the word “noun” is used to discuss about syntax, and the words “request”, “female,” “restaurant” are used to discuss about language in relation to gender, politeness, or setting. Linguistics explains the jokes, sayings, idioms, songs of everyday culture and study about these into pieces as how

language works. Linguistics more or less studies the small *c* of speech, while Cultural Studies study the capital *C* in which more complex topics should be discussed. It is proven by the way Linguistics and Cultural Studies use general words in their study. For example, Linguistics tends to use the words “grammar,” “translation”, and “spelling” since it studies about the language structures and application in daily life. However, Cultural Studies and Literature have their own specific words according to the fields. These sub-genres do not really differ in terms of GSL use. For example, Literature and Cultural Studies cover general words, such as “commercial,” “power,” and “represent.” Even though Linguistic also studies about the use of language in relation to society and culture, the topics are not wide. For example, in sociolinguistics, we study about lexical choices in female or male speech and relate it to gender concept. The context discussed in Linguistics involves certain issue and does not relate to many more aspects. Linguistics indeed scientifically analyzes the language, such as its structure, history and relations between languages, the use of language in social context, language acquisition, and many more. The most important thing to be emphasized here is Linguistics still has relation to culture, but the scope of its topics is narrower than Cultural Studies. Therefore, Linguistics has more general words since it extracts the use of language in everyday life and analyzes it scientifically. From Table 3, level 1 and 2 show Linguistics has less types than Cultural Studies and Literature. This finding opposes the opinion of Akbarian, Ghanbarzadeh, & Shahri which says more word types and word families found in a text infer that the writer has more vocabulary proficiency and competence in choosing and using lexicon [9].

TABLE IV. TEN TOP LIST OF GSL IN LEVEL

Linguistics		Literature		Cultural Studies	
noun	14	aims	6	fans	13
request	9	resistance	6	advertisement	11
female	7	island	5	consciousness	7
females	6	connection	3	film	6
males	6	dancing	3	aims	5
bargaining	5	examines	3	female	5
restaurant	5	wicked	3	during	4
shopping	5	absence	2	especially	4
aims	4	encouragement	2	examine	4
during	4	examine	2	commercial	3

Literature abstract have less general words than Linguistics. What we can infer about the result is that Literature basically has some similarities to Cultural Studies. Literature also studies some theories that Cultural Studies study, such as Marxism, Structuralism, and Power Relation. However, what makes them different in relation to the use of general words is the way it is presented. Literature is always related to literary works, such as poems and novels, but it is

written in the way literary work works. The way Literature presents their analysis is by narrating and analyzing the story. Therefore, the words are not extremely academic because the narration usually uses general words to reflect more human’s life. The narrative stories in the written text become as the only source of Literature study. However, the issue also can cover parts of in Cultural Studies. Regarding to the source of data, the data are gathered from the written text itself and in its characters’ point of view. The combination between story and cultural phenomena embodied in literary texts influence the way authors present their ideas in abstracts. As mentioned previously, the text in Literature study becomes the only data for analysis process. The text, of course, contains more or less common words that people understand when they read them since the literary work aims for entertainment. The words “soldiers”, “stories”, “experience”, “summer” are the examples of words which we are familiar with in daily life.

What has been talked previously is in line with the finding of Akbarian, Ghanbarzadeh, & Shahri [10]. They asserted that the different objectives and conceptualizations of vocabulary influence to the richness of the arguments on vocabulary.

C. Academic Word List Profiles of Linguistics, Cultural Studies, and Literature

The academic word list profiles of the three domains of interest shows that Cultural Studies undergraduate thesis abstracts contain the highest percentage of words from the AWL, with a total percentage of 12.4% compared to 9.2% for Linguistics and 9.1% for Literature. Cultural Studies abstracts also lead in terms of total AWL tokens and types, with 300 and 162 respectively. This is particularly impressive given the fact that the total number of types overall is 287, which means that CS included more than 50% of the total academic types. In comparison, The Linguistics abstracts showed that they had 224 tokens and 107 types whilst their Literature equivalents were 198 and 110 respectively. The Cultural Studies abstracts also held the most number of groups (118) compared to Linguistics and Literature (both 84). Overall, the top 10 words in AWL of the Cultural Studies abstracts had more usage than both its Linguistics and Literature counterparts.

TABLE V. NUMBER OF AWL TOKENS, GROUPS, AND TYPES IN EVERY SAMPLE GROUP

Field	Tokens	Type	Group
Linguistics	224	107	84
Literature	198	110	84
Cultural Studies	300	162	118
Overall	722	287	194

TABLE VI. TOP 10 AWL WORDS IN EVERY SAMPLE GROUP ALONG WITH FREQUENCY

Linguistics		Cultural Studies		Literature	
Type	Freq	Type	Freq	Type	Freq
Feature	11	Gender	14	Theory	9
Data	9	Media	11	Qualitative	8
Attitude	8	Culture	8	Adults	7
Method	6	Theory	8	Research	7
Policy	6	Analysis	7	Method	6
Qualitative	6	Cultural	6	Analysis	5
Research	6	Data	6	Cultural	5
Analyzed	5	Qualitative	6	Analyzze	4
Approach	5	Ideology	5	Construction	4
Interaction	5	Issue	5	Text	4

This result showcases the high amount of academic words used in cultural studies and low amount in literature, but raises the question of how and why this is the case. In particular, the large gap compared to both literature and linguistics is interesting. In addition, the similarity between linguistic and literature academic word profiles is equally interesting. To answer those questions, we must first compare the word profiles in pairs and find evidence to reasons for this occurrence.

We will first compare the AWL profile of cultural studies and linguistics. Cultural studies leads linguistics in all of the parameters of academic word usage, underlined by the difference of 3.2 percentage points (12.4% compared to 9.2%). This is largely attributed to the high usage of common words (GSL) in linguistic abstracts (a total of 79.2% compared to 72.2% in the case of cultural studies).

There are a few possible reasons why this is the case. The first reason is the possible difference in object. Linguistic research often involves vocabulary and registers which includes many common words. This means that GSL words can easily be the object of research for linguistic papers. For instance, one of the abstracts features a study on nouns relating to the word “beautiful,” including words such as “woman,” “body,” “place,” and so on. Most of these are common words. For studies in fields like second language acquisition too, it is not hard to imagine that most words used to describe the study will repeat common key words, such as “teacher” and/or “learning” (thus being GSL).

In contrast, the object of cultural studies are usually described using words regarded as academic, for instance “gender” (which is the most frequently used word in the cultural studies abstract sample), “media,” “culture,” and so on. This phenomenon is proven by the fact that the 5 most frequent words in the CS (cultural studies) vocabulary profile is used more frequently than the 5 most frequent words in the Linguistics vocabulary profile. Objects of a study would probably be mentioned several times in an abstract, and thus if objects and words relating to objects in CS are described using academic words and unlisted words (which will be discussed

later on) while objects of linguistics in and of themselves are already common words and further have many related common words, then it may be a factor in this gap of academic words.

The second probable cause is scope. This is also related to the object of the two areas of research. Linguistics covers anything relating with language, thus it will be more likely to include common words, mostly within GSL, as previously mentioned. However, because linguistics specializes in language, most specialized words in linguistics refer to concepts and registers that are mostly used either within linguistics or in other specialized fields in different contexts. This means that the usage of words describing concepts in linguistics such as “syntax” and “morpheme” are more likely to be classified as technical or uncommon rather than academic, because they are not used often in the broader range of academic works.

TABLE VII. COMPARISON BETWEEN TOP 10 MOST USED ACADEMIC WORDS BETWEEN LINGUISTICS AND CULTURAL STUDIES

Linguistics		Cultural Studies	
Type	Freq	Type	Freq
Features	11	Gender	14
Data	9	Media	11
Attitude	8	Culture	8
Method	6	Theory	8
Policy	6	Analysis	7
Qualitative	6	Cultural	6
Research	6	Data	6
Analyzed	5	Qualitative	6
Approach	5	Ideology	5
Interaction	5	Issue	5

Meanwhile CS covers a wide range of social issues which often requires language that are more widely used in the academic community, thus becoming common academic words instead of specialized ones. For example, the term “gender” is used in all three fields and is listed in the AWL, and it is widely used in the CS abstracts. This contrasts with linguistic core terms which, as discussed, are more specialized. For example, a sociolinguist, a sociologist, a biologist and a psychologist might use the word gender, thus the word becomes an academic word because it is often used in an academic setting. However the chances of someone outside of Linguistics, Mathematics and Computer Science using the word “syntax” is much lower, resulting in it having less of a chance to be classified as an academic word. This is because the AWL accounts for words used often in academia, not technical ones.

Furthermore, since many of the concepts in CS are described in these kinds of “commonly used technical terms”, CS abstracts are more likely to utilize a wider range of academic words in order to cover its research topics, whilst

linguistics rely on a more specialized mix of an array of academic words and technical words.

This is indicated by the difference in types among the two vocabulary profiles, where CS is more than Linguistics. In essence, CS abstracts are more likely to include more academic words because many of their core terms such as “gender”, “culture”, and so on are widely used in other fields, and is thus listed in the AWL, while linguistics use a mix of common academic words and their own core terms like “adjective” and “dialects”, listed in dictionaries such as *The Concise Oxford Dictionary of Linguistics* which are less likely to be classified in the AWL [11].

The comparison between the AWL vocabulary profiles of CS and Literature is a different matter. At first glance, it can be easy to assume that both would have similar vocabulary profiles, given they share many common intersecting topics and objects including gender, social constructs, media, and so on. Indeed, they have similarities such as the fact that both Literature and CS have more unlisted words than Linguistics. However, it is also clear that there CS has more academic words than Literature; in fact the gap is slightly wider than it is with Linguistics at 3.3%. As with CS and Linguistics, there are several possible reasons for this gap.

One reason is that Literature research abstracts simply are briefer than CS or Linguistics abstracts in this study. This is proven by the fact that the total tokens of the sample CS and Linguistics abstracts are 2419 and 2442 respectively, whilst the total of Literature sample abstracts’ tokens is 2164; a difference of 200 words or more with each other sample group. This is significant, considering despite the lesser amount, the Literature dataset provided comparable numbers in AWL tokens, percentage, and even beat Linguistics on amount of types (198, 9.1%, and 110, respectively compared to 224, 9.2% and 107 in Linguistics). This indicates that Literature research abstracts are similar to CS compared to Linguistics in the sense that they both tend to use academic words (and uncommon words) more often than common words in general. It could be concluded that if Literature abstracts were not as brief, perhaps more academic words may be used to make it comparable to CS.

TABLE VIII. LIST OF TOP 10 UNLISTED WORDS IN LINGUISTIC ABSTRACTS

Type*	Range	Frequency
Lexical	1	10
Adjective	1	8
Classroom	1	8
Dialect	1	7
Feedback	1	6
Indonesian	1	6
Malls	1	5
Simpsons	1	5
Sub	1	5
Adjectives	1	4

*Note that several words commonly used in linguistic academia such adjectives, lexical, and dialect are presented and widely used.

TABLE IX. TOTAL TOKENS, TYPES, AND GROUPS OF ALL THREE DATASETS

No	Field	Total Tokens	Total Type	Total Groups
1	Linguistics	2442	701	560
2	Cultural Studies	2419	790	639
3	Literature	2164	691	576

Another reason is of course, the fact that Literature abstracts seem to use more uncommon or unlisted words compared to CS, which accounts for some of the gap in the AWL profile (16.1% compared to 15.4%). The most notable part here is of course that despite having more than 200 tokens less in its profile, Literature abstracts managed to have a comparable amount of unlisted tokens (348) compared to CS (372).

This simply means that Literature abstracts are more likely to use unlisted words rather than academic words compared to CS. This can be attributed to the fact that Literature abstracts tend to mention names and uncommon terms used within a certain work compared to CS abstracts.

While CS abstracts utilize academic words such as gender and media very frequently, Literature abstracts are more prone to mentioning names; presumably of authors, characters, and people of interest concerning their topic. This is proven by the fact that two of the five most frequent unlisted words in the Literature vocabulary profile are names (Anna and Malala). Like how Linguistics abstracts often mention their objects which usually necessitate mentioning common words, Literature abstracts seem to require mentioning their objects and describing them with uncommon words. Although this is expected from such a field, it is also true that this means that it tends to use less academic words than CS.

TABLE X. TOP TEN AWL IN LITERATURE

Type	Range	Frequency
Novel	1	19
Anna	1	9
Dysfunctional	1	8
Imperialism	1	8
Malala	1	8
Ecological	1	7
Homer	1	6
Iliad	1	6
Oppression	1	6

Of course, equally interesting is the similarities between the AWL profiles of the Linguistics abstracts and the Literature abstracts. It has been mentioned before, but both sets of data offer a near exact same percentage of academic words (9.2% in Linguistics compared to 9.1% in Literature). Given the difference in total tokens, it is not strange that the Linguistic data set has more academic word tokens total compared to Literature (224 compared to 198). The interesting part is how Literature abstracts actually have more academic word types than Linguistics (110 compared to 107). The difference is small, but considering the gap in total tokens, it is nonetheless notable. This may be attributed to Literature’s similar nature to CS compared to Linguistics. As previously mentioned, many

Linguistic terms use more common words to describe their objects because their objects are often those words themselves whilst CS and Literature seem to use more academic words relating to society and more widely used in wider academics. As with CS and linguistics, perhaps the scope of the research is a factor as well. Just as linguistics specializes in language in many contexts, literature specializes in its namesake. This might mean that both linguistics and literature share a similarity in which they both do not have the breadth of scope that CS covers, thus even if they use academic words more like CS than Linguistics, they are still using more specialized words than CS. In conclusion, CS abstracts make use of the most academic words, both in tokens and types, because they use them in both describing their objects and the theories they use. Linguistics on the other hand relies more heavily on specific terms for theories and common words for describing their objects and Literature abstracts tend to be similar in how they use academic words to CS compared to with Linguistics, but have overall an lower average number of total tokens which affects their academic word count. They also use mostly unlisted words to describe their objects, instead of academic words.

D. Low-Frequency and Technical Words

Among the three domains of interest in this study, the undergraduate abstracts from Literature shows the highest percentage of unlisted words with 16.10% coverage. It is then followed by the abstracts from Cultural Studies with 15.40% coverage and Linguistics with 11.70% coverage. It is important to take notice that unlisted word list is established by low-frequency words and technical words in the subject area. Low-frequency words itself refer to the words which are not included in general words, academic words, and technical words.⁵ Words which usually fall into this type is words which are rarely used in a language and proper names such as names of people, countries, and cities. Meanwhile, technical words refer to the words which are common in a specific topic or subject area only because they belong to the particular topic or subject area. Below is a table showing the unlisted words in top five most-used rank based on the frequency of usage in the abstracts of each major.

The fact that Literature abstracts containing the highest percentage of unlisted words among all is actually predictable. The reason is that the Literature abstracts which we used as our data generally deal with issues in book(s) in which the names of the characters are mainly discussed. This statement is hugely supported by the finding that “novel” is the predominant word in top five unlisted word list for Literature abstracts. As explained in the earlier explanation, proper names are the main contributors of low-frequency words. This is reflected on the data that five out of fourteen most-used words on the list are “anna”, “malala”, “homer”, “iliad”, and “tolstoy”. Anna is the name of a character in a book; Malala, Homer, and Tolstoy are the names of book writers; while Iliad is the title of a poem. As for the technical words, there are literary terms from the list which can be found in *The Concise Oxford Dictionary of Literary Terms*, namely “novel” and “deconstruction” [12].

The same goes with Cultural Studies abstracts which are ranked the second highest percentage of unlisted words. As

widely known, Cultural Studies in nature deal with social or cultural phenomenon in specific areas or in media such as films, advertisements, etc. It is clearly reflected through the most-used unlisted words in the Literature abstracts which are “Indonesia”, “nz” which stands for New Zealand, and “Bali” which refer to a specific place and “online”, “video”, and “Instagram” which are related to media. On the list, we can also find three technical terms in Cultural Studies which are “masculinity”, “discourse”, and “nationalism”. The first two terms are listed in *The Sage Dictionary of Cultural Studies*¹⁴ and the last one is listed in *A Dictionary of Cultural and Critical Theory* [13]. Even so, the term “discourse” is not a specific technical term only for Cultural Studies because it is also used in Linguistics field.

Meanwhile, Linguistics abstracts have the least percentage of low-frequency words presumably because Linguistics theses generally do not use proper names as often as Literature and Cultural Studies theses. It is reflected on the top most-used unlisted words list that only two out of nine words in Linguistics which are names. It is presumably because the discussion of Linguistic studies revolves around the topic about language. This is supported by the second-ranked most-used unlisted words on the Linguistics list which are “lexical” and “adjective” which are specific terms in language learning. Both are included as technical terms that can be found in *Thse Concise Oxford Dictionary of Linguistics* [14].

IV. CONCLUSION

In conclusion, based on our data and analysis, Linguistic abstracts have the most number of GSL words and least number of unlisted words, Cultural Studies have the most AWL words but least percentage of common (GSL) words while Literature has the most unlisted words but least amount of total tokens, GSL words and AWL words. These differences in profiles reflect the different use of the different types of words in the three fields. Linguistics cover language itself and thus use GSL not only as common words but also especially to describe their objects, and their objects are sometimes the GSL words themselves. Meanwhile, Cultural Studies use mostly academic words in their broad field, often describing objects with words commonly used by other parts of academia more so than Literature and Linguistics. Literature is similar to Cultural Studies in its makeup, but uses comparatively far more unlisted words due to its objects necessitating it. Linguistics use unlisted technical words for describing concepts like “adjective” and “verb” more so than names while Cultural Studies use them similarly to Literature, but not nearly as much. These results prove that even within one department there can be interesting insights into the differences in how different levels of words are used within the different sub-fields.

REFERENCES

- [1] Plunkett, K., Lexical Segmentation and Vocabulary Growth in Early Language Acquisition. *Journal of Child Language* (Vol. 20, No.1), pp.43-60, 1993.

- [2] Morris, L., & Cobb, T., Vocabulary Profiles as Predictors of the Academic Performance of Teaching English as a Second Language Trainees. *System* (Vol. 32, No. 1), pp.75-87, 2004.
- [3] Valipouri, L., & Nassaji, H., A Corpus-based Study of Academic Vocabulary in Chemistry Research Articles. *Journal of English for Academic Purposes* (Vol. 12, Issue 4), pp.248-263, pp.201, 2013.
- [4] Martínez, I. A., Beck, S. C., & Panza, C. B., Academic Vocabulary in Agriculture Research Articles: A Corpus-based Study. *English for Specific Purposes* 28, pp.183-198, 2009.
- [5] Nation, I. S., *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press, 2001.
- [6] Mackey, Alison. & Gass, Susan., *Second Language Research: Methodology and Design*. New Jersey: M. LEA, 2005.
- [7] Creswell, J.W., *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches*. The 3rd Ed. UK: SAGE, 2009.
- [8] West, M., *A General Service List of English Words*. London, England: Longman, 1953.
- [9] Akbarian, Is'haaq., Ghanbarzadeh, Zahra., & Shahri, Mohammad A., Profiling Vocabulary in Psychology Journal Abstracts: A Comparison between Iranian and Anglo-American Journals. *Iranian Journal of Language Teaching Research*, 5(1), pp.51-69, 2017.
- [10] Akbarian, Is'haaq., Ghanbarzadeh, Zahra., & Shahri, Mohammad A., Profiling Vocabulary in Psychology Journal Abstracts: A Comparison between Iranian and Anglo-American Journals. *Iranian Journal of Language Teaching Research*, 5(1), pp.51-69, 2017.
- [11] Baldick, C., *The Concise Oxford Dictionary of Literary Terms*. Oxford: Oxford University Press, 2008.
- [12] Baldick, C., *The Concise Oxford Dictionary of Literary Terms*. Oxford: Oxford University Press, 2008.
- [13] Barker, C., *The Sage Dictionary of Cultural Studies*. London: SAGE Publications Ltd, 2004.
- [14] Matthews, P. H., *The Concise Oxford Dictionary of Linguistic (Second Edition)*. Oxford: Oxford University Press, 2007.