Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi
Ondokuz Mayıs University Journal of Faculty of Education
e-ISSN: 2548-0278 OMU EFD, December 2022, 41(2): 617-638

The Role of Affixation on Lexical Richness of EFL Learners' Written Texts

Ek İle Kelime Türetmenin İngilizceyi Yabancı Dil Olarak Öğrenen Öğrencilerin Yazılı Metinlerindeki Sözcüksel Zenginlik Üzerindeki Rolü

Mustafa YILDIZ[1]

[1]Samsun Üniversitesi, Samsun
• myildiz55@yahoo.com • ORCİD > 0000-0002-7971-5847

Makale Bilgisi / Article Information

Makale Türü / Article Types: Araştırma Makalesi / Research Article
Geliş Tarihi / Received: 05 Eylül / September 2022
Kabul Tarihi / Accepted: 10 Aralık / December 2022
Yıl / Year: 2022 | **Cilt – Volume:** 41 | **Sayı – Issue:** 2 |**Sayfa / Pages:** 617-638

Atıf/Cite as: Yıldız, M. "The Role of Affixation on Lexical Richness of EFL Learners' Written Texts" Ondokuz Mayıs University Journal of Faculty of Education, 41(2), December 2022: 617-638.

https://doi.org/10.7822/omuefd.1171099

THE ROLE OF AFFIXATION ON LEXICAL RICHNESS OF EFL LEARNERS' WRITTEN TEXTS

ABSTRACT:

The current study explores EFL learners' productive affix knowledge in various essay types. More specifically, the study investigates whether the essay types differentiate from each other with regard to the production of words that have no affixes, inflected words, and derived words included in the first and second 1,000 word frequency bands. It further examines both the most frequent affixes and the distribution of affixes mentioned in Bauer and Nation's framework (1993) across essay types. A group of datasets for each of four different essay types have been formed. In each of these datasets, firstly, the distribution of words across the first 1,000 and second 1,000 word frequency bands was investigated. Secondly, both basewords and affixed words are analyzed in each of those two word frequency bands. Thirdly, the distribution of affixes across Bauer and Nation's framework (1993) is investigated. Lastly, the most frequent affixes in each of the datasets are explored. The findings reveal that the essay type plays a partly active role in use of words included in different word frequency bands. Additionally, as the word frequency band decreases, the number of affixed words increases. Also, affix types mentioned in Bauer and Nation's framework (1993) do not reflect the level of difficulty according to the findings obtained from the existing datasets.

Keywords: *Affixation, Affix, Morphology, Prefix, Suffix.*

❋❋❋

EK İLE KELİME TÜRETMENİN İNGİLİZCEYİ YABANCI DİL OLARAK ÖĞRENEN ÖĞRENCİLERİN YAZILI METİNLERİNDEKİ SÖZCÜKSEL ZENGİNLİK ÜZERİNDEKİ ROLÜ

ÖZ:

Mevcut çalışma, İngilizceyi yabancı dil olarak öğrenen öğrencilerin çeşitli kompozisyon türlerinde üretken ek bilgilerini araştırmaktadır. Daha spesifik olarak mevcut çalışma, birinci ve ikinci 1,000 kelimelik sıklık bandında yer alan ek almamış kelimeler, çekim eki almış kelimeler ve türemiş kelimelerin üretimi açısından kompozisyon türlerinin birbirinden farklılaşıp farklılaşmadığını araştırmaktadır. Ayrıca Bauer ve Nation'ın çalışmasında (1993) bahsedilen ekler arasında hem en sık kullanılan ekleri hem de bu eklerin kompozisyon türleri arasındaki dağılımını incelemektedir. Dört farklı kompozisyon türünün her biri için bir grup veri seti oluşturulmuştur. Bu veri setlerinin her birinde, ilk olarak, kelimelerin ilk 1,000 ve

ikinci 1,000 kelime frekans bandındaki dağılımı incelenmiştir. İkinci olarak, hem ek almamış kelimeler hem de ek almış kelimeler bu iki kelime frekans bandının her birinde ayrı ayrı analiz edilmiştir. Üçüncü olarak, Bauer ve Nation'ın (1993) çalışmasında yer alan eklerin dağılımı araştırılmıştır. Son olarak, veri setlerinin her birinde en sık kullanılan ekler incelenmiştir. Bulgular, kompozisyon türünün farklı kelime frekans bantlarında yer alan kelimelerin kullanımında kısmen etkin bir rol oynadığını ortaya koymaktadır. Ayrıca kelimelerin dahil oldukları frekans bandı azaldıkça ek almış kelime sayısı da artmaktadır. Ayrıca, Bauer ve Nation'ın (1993) çalışmasında bahsedilen ek türleri, mevcut veri setlerinden elde edilen bulgulara göre zorluk derecesini yansıtmamaktadır.

***Anahtar Sözcükler**: Ekleştirme, Ek, Biçimbilim, Önek, Sonek.*

❋❋❋

INTRODUCTION

The statement 'I know this word' is highly prone to be interrogated and can raise many questions about different aspects of word knowledge. Based on the multidimensionality of word knowledge Nation (2013) suggests, the very general questions people come across as follows interrogate different aspects of someone's word knowledge: Is the word a person claims to know a part of receptive or productive vocabulary? Is the structure of the word claimed to be known recognized and produced in both spoken and written language? Does the word make sense in mind; and can it be matched with a referent meaningfully? Is the word recognizable in a grammatical/lexical pattern; and can it be used in a suitable linguistic context? All these questions refer to the multidimensionality of word knowledge, which makes any possibility impossible except the gradual progress of word learning.

Internal morphological structure of words, mentioned as word parts within the multidimensionality of word knowledge (Nation, 2013), is also one of the indispensable aspects of vocabulary knowledge. The recognition of the most meaningful parts within the structure of a word and the use of these morphological items to convey meaning stand out as a different dimension of word knowledge. Nation (2020) emphasizes the gradual progress of word part knowledge by underlining that "knowledge of word parts primarily involves being able to use the inflectional system of the language, with the next step involving the more gradual growth of knowledge of the derivational affixes" (p. 18), suggesting that derivational morphology has the potential of much more learning burden for learners compared to inflectional morphology.

Affixation in the word formation process has been handled from many different perspectives such as the contribution of derivatives to reading comprehension

(Laufer & Cobb, 2020); the relationship between learners' vocabulary size and their affix knowledge (Danilović et al., 2013; Hayashi & Murphy, 2011; Mochizuki & Aizawa, 2000); the acquisition order of affixes (Danilović et al., 2013; Mochizuki & Aizawa, 2000; Tamura & Shirahata, 2016); and the validity of difficulty ranking of affixes suggested by Bauer and Nation (1993) (Leontjev, 2016; Tamura & Shirahata, 2017). Among these aforementioned studies, Laufer and Cobb (2020) investigate a group of texts to analyze the distribution of words with prefixes and suffixes and their contribution to the lexical coverage of the texts required for minimal and optimal levels of reading comprehension. A variety of texts such as applied linguistics articles, news articles, simplified and unsimplified novels are analyzed in terms of the number of basewords without affixes, words with inflections, and words created at the end of the affixation process and classified according to each level mentioned in Bauer and Nation (1993). They find that the percentage of basewords within these various text types disperses between 75% and 87%. Inflected words increase this rate to around 94%. The remaining 6% consists of derived words. While 26 different affixes are used to form derived forms in newspaper articles, this number drops to 7 in simplified reading texts. In addition, while 50% of the total derived words are formed by only three suffixes; -ly, -ion, -er, a total of 17 different suffixes are involved in the formation of more than 95% of the derived words.

Mochizuki and Aizawa (2000) search for the relationship between the size of learners' vocabulary and affix knowledge. They further investigate the acquisition order of affixes. The results reveal that an average number of 3,769 words are known to the participants. Additionally, more than half of both 13 prefixes and 16 suffixes, on average, are known to the participants. As they increase their vocabulary, their affix knowledge necessarily increases. Mochizuki and Aizawa (2000) also assume that an affix which is known by a large number of learners means that it is acquired earlier than a lesser-known affix. Re-, un-, pre- are the prefixes that the participants decipher the meaning most easily while -ation, -ful, -ment are the suffixes whose meanings they decipher the most easily. Similarly, Danilović et al. (2013) search for the interrelatedness between Serbian upper-intermediate EFL learners' affix knowledge and their receptive and productive vocabulary size. They (2013) also investigate the acquisition order of affixes for Serbian EFL learners. The findings suggest that while a moderate positive correlation is available between the overall receptive vocabulary and prefix knowledge, any statistical correlation is identified between the overall receptive vocabulary and suffix knowledge. In a similar vein, the findings reveal that Serbian EFL learners' productive vocabulary and their prefix knowledge significantly correlate with each other; however, no correlation between their productive vocabulary and suffix knowledge is identified. Danilović et al. (2013) conclude that the prefixes post-, anti-, and re- are the affixes which are acquired earlier because they are the most accurate prefixes known to Serbian EFL learners.

Tamura and Shirahata (2016) conduct a study similar to Mochizuki and Aizawa (2000) with partial changes in the method of the study. Mochizuki and Aizawa (2000) test Japanese EFL learners' prefix knowledge by utilizing non-existing words created with the use of existing prefixes to eliminate the positive effects of the knowledge of already known affixed words on the prefixes to be tested. In order to test whether the prefix order suggested by Mochizuki and Aizawa (2000) is valid or not, Tamura and Shirahata (2016) investigate Japanese EFL learners' prefix knowledge by using existing derivatives created by means of prefixes being tested. The findings reveal that the prefixes non-, semi-, pre-, re-, and anti- are the prefixes which are highly accurate. The prefix ante- is highly difficult at the end of the prefix ranking with the accuracy level of 18.5%. Tamura and Shirahata (2016) and Mochizuki and Aizawa (2000) are highly correlated with each other, suggesting that prefixes having been tested might have really a fixed order of difficulty for Japanese EFL learners.

Hayashi and Murphy (2011) compare Japanese ESL learners with native speakers of English regarding receptive and productive knowledge of inflections and derivations. As for receptive measure of affix knowledge, they found that Japanese ESL learners significantly score better on derivative affixes which change the grammatical class of words than native speakers of English do. However, the groups do not differ in terms of their performance on the derivative affixes which do not change the part of speech of the words. On the other hand, as for productive measure of affix knowledge, native speakers of English achieve a very high accuracy rate in the production of inflectional suffixes and class-changing and class-maintaining affixes. Hayashi and Murphy (2011) also suggest that Japanese ESL learners' productive derivational knowledge positively correlates with their receptive and productive vocabulary size.

Tamura and Shirahata (2017) explore if Bauer and Nation's (1993) ranking of prefixes is valid for Japanese adult learners of English. More specifically, Tamura and Shirahata (2017) investigate the extent to which the order of difficulty of prefixes for Japanese students overlaps with the one mentioned by Bauer and Nation (1993). At the end of the assessment of Japanese learners' English prefix knowledge, Tamura and Shirahata (2017) conclude that although their findings partially overlap with Bauer and Nation's (1993) affix difficulty ranking, they fundamentally differentiate from each other. Non-, hyper-, and semi- are three prefixes with over 90% accuracy. On the other hand, the prefixes ante-, circum-, and sub- reach 10% accuracy level. Although the prefixes semi- and hyper- are considered to be difficult in Bauer and Nation's (1993) affix difficulty ranking, they are among the easiest prefixes in Tamura and Shirahata (2017). Similar to Tamura and Shirahata (2017), Leontjev (2016) investigates to what extent the difficulty learners experience with the recognition of derivational affixes differs significantly from the affixes mentioned by Bauer and Nation (1993). Leontjev (2016) finds that while the mean

number of affixes recognized at Levels 5 and 6 does not significantly differ, the number of affixes at the other levels (3-4) dramatically differs. As the level of affixes increases, there is a significant decrease in the number of recognized affixes. About 75% of the affixes at Level 3 are recognized by the participants, while about 50% of the affixes at Level 4 are recognized. For the affixes at Levels 5 and 6 this recognition rate drops to around 25%. Contrary to Tamura and Shirahata (2017), the findings revealed by Leontjev (2016) largely support the affix order mentioned by Bauer and Nation (1993).

Examining 100 million word-sized data from the British National Corpus, Nation (2013) finds that the most frequently used first 2,000 words constitute 86.06% of the corpus, referring to the vital importance of high frequency words to learning a language and building vocabulary. Nation (2013) also underlines the importance of the number of members available in a word family. While the number of word family members in the first 1,000 word frequency band in English is 6.8 on average, this mean value drops to an average of 1.8 for the twentieth 1,000 word frequency band. To visualize better, if a student who knows the headwords in the first 1,000 word frequency band also knows other word family members formed by inflectional and derivational affixes belonging to these headwords, it necessarily means s/he knows 6,838 individual words on average. Assuming that the same learner knows all the word families with all of the word family members in the second 1,000 word frequency band, it is needed to add an additional 6,367 words to the number of individual words that the learner knows. As the frequency band in which a word family is included decreases, the number of words that are members of this word family also decreases. That is, when the same learner learns the most frequently used twentieth 1,000 word family with all its members, s/he will add 1,810 additional words to his/her vocabulary, which implicitly refers to the importance of affixation, especially for high-frequency words, generating word family members for headwords through either inflectional or derivational affixes.

The following quotation from Nation (2013), "(t)he number of words in a word family is very strongly related to family frequency. The more members a family contains, the more frequent it is likely to be" (p. 393), constitutes the arising point of the present study. As mentioned in the former paragraph, Nation (2013) asserts that the more members a word family has, the more likely that word family is among the most frequent word frequency bands. The present study is conducted to test Nation's (2013) this claim. It explores whether the most frequently used word families actually have the highest number of word family members in EFL learners' productive vocabulary. The study also investigates the distribution of affixes based on Bauer and Nation's (1993) classification. More specifically, the following research questions will be tried to be answered:

1. What is the distribution of the words included in the first and second 1,000 word frequency bands across the essay types?

2. Do the essay types significantly differ from each other with regard to EFL learners' use of basewords, inflected words and derived words included in the first and second 1,000 word frequency bands?

3. What is the distribution of affixes mentioned in Bauer and Nation's framework (1993) in each of the essay types?

4. What are the most frequently used derivative affixes in each of the essay types?

METHOD

Data

The data of the current study consists of 71 essays randomly selected from the essays written by EFL learners at C1 level of English proficiency. A dataset of 40,000 words was created from four different essay types because the system Morpholex Affix Profiler (Cobb, accessed 01 June 2022), used to analyze the data of the study in terms of Bauer and Nation's affix framework (1993), allows the analysis of 10,000 words at a time. When the ten-thousand-word limit was reached, the dataset of each essay type was completed and no more essays were added. The datasets consisting of argumentative and comparison-contrast essays contain 19 essays for each, while the dataset consisting of cause-effect essays contains 17 essays, and the dataset consisting of descriptive essays contains 16 essays.

Bauer and Nation (1993) create a hierarchy of affixes by classifying them in terms of frequency, regularity, predictability and productivity. The hierarchy contains 7 levels of affixes, one level of which is a group of inflectional affixes while five levels of which contain five different groups of derivational affixes. The hierarchy also contains a group of roots coming from Greek or Latin which are not included in this study as the items in this group do not serve for the purpose of the present study. More specifically, Level 1 consists of basewords which do not have any prefixes or suffixes, meaning that each individual lexeme is an individual word. Level 2 contains a group of inflectional suffixes available in English. Level 3 is formed of the most frequent and regular derivational affixes. Level 4 contains frequent and regular affixes. While infrequent but regular affixes are grouped as Level 5, frequent but irregular affixes are grouped as Level 6 affixes.

Data Analysis

The first research question investigates whether there is a significant difference among four datasets in terms of EFL learners' productive vocabulary which belong to the first and second 1,000 word frequency bands. First, to determine the number of words in each of the word frequency bands, Compleat Web VP System (Cobb, accessed 01 June 2022) has been used. The system classifies the words in each dataset based on the frequencies of the words in the BNC/COCA corpora. After sorting out the words of the first and second 1,000 word frequency bands, each of the datasets was compared to each other by applying the test of Multivariate analysis of variance (MANOVA) to reveal whether they differentiate from each other in terms of the use of words of these two word frequency bands.

The second research question explores whether the datasets significantly differentiate from each other regarding EFL learners' use of basewords, inflected words and derived words included in the first and second 1,000 word frequency bands. The Morpholex Affix Profiler System (Cobb, accessed 01 June 2022) classifies the words included in the first and second 1,000 word frequency bands as basewords, inflected words, and derived words. Each of the four datasets was first analyzed according to the morphological structure of the words in its content and then compared with each other by using the test of Multivariate analysis of variance (MANOVA) to see whether the datasets significantly differ from each other in terms of the use of basewords, inflected words, and derived words of the first and second 1,000 word frequency bands.

The third research question investigates the distribution of affixes mentioned in Bauer and Nation's framework (1993) in each of the datasets. To answer this research question, on the one hand, the distribution of basewords, inflected words, and derived words in each of the overall datasets was determined. On the other hand, the affixes are classified according to Bauer and Nation's framework (1993).

The fourth research question tries to identify the most frequently used derivative affixes in each of the datasets. The Morpholex Affix Profiler System (Cobb, accessed 01 June 2022) helps analyze the derivative affixes in each of the datasets according to their frequency.

FINDINGS

The present study was conducted to examine EFL learners' productive affix knowledge in their written texts. More specifically, the study was carried out to determine whether the essay types differ in terms of the production of words that have no affixes, inflected words, and derived words included in the first and second

1,000 word frequency bands. The distribution of affixes mentioned in Bauer and Nation's framework (1993) across essay types and the most frequent affixes in each of these different essay types were further examined.

The first fifteen tables below are included to report the answers to the first and second research questions. Tables 16-19 report the answers to the third research question while the last table (Table 20) contains the answer to the last research question.

Table 1. Descriptive Statistics for the Distribution of Words Included in First and Second 1,000 Word Frequency Bands across Datasets

	Essay Types	Mean	SD
1K Tokens	Argumentative	77,8632	3,67593
	Descriptive	86,3500	3,24448
	Comparison-Contrast	75,0421	3,30123
	Cause-Effect	73,5294	4,24540
	Total	77,9831	5,98517
2K Tokens	Argumentative	10,8579	2,40447
	Descriptive	7,3125	2,51048
	Comparison-Contrast	12,1737	2,09864
	Cause-Effect	16,8000	3,92205
	Total	11,8338	4,29391

The findings indicate that descriptive essays consist of less challenging words. The ratio of the words from the most frequently used two thousand words to the total number of words in descriptive essays is almost 94%. The essay type, in which the number of words that are less demanding is the least, is the comparison-contrast essay. Cause-effect essays stand out with the highest number of words included in the second 1,000 word group while descriptive essays do with the least number of words included in the same word frequency band.

Table 2. MANOVA Test Results for Words Included in First 1,000 Word Frequency Band

Word Frequency Band	Essay Types	Essay Types	Mean Difference	Std. Error	Sig.
1K Tokens	**Argumentative**	Descriptive	-8,4868	1,23366	,000
		Cause-Effect	4,3337	1,21380	,004
	Descriptive	Argumentative	8,4868	1,23366	,000
		Comparison	11,3079	1,23366	,000
		Cause-Effect	12,8206	1,26640	,000

It seems that the categorical independent variable, i.e. essay type, affects the participants' productive vocabulary among the first 1,000 words. With regard to the use of the first 1,000 words, it is apparent that the number of words used in the descriptive essay type which are among the first 1,000 words differs significantly from the number of words in the same word frequency band used in the other three essay types. The mean difference of the productive use of the first 1,000 words in the argumentative and cause-effect essay types is also statistically significant.

Table 3. MANOVA Test Results for Words Included in Second 1,000 Word Frequency Band

Word Frequency Band	Essay Types	Essay Types	Mean Difference	Std. Error	Sig.
2K Tokens	**Descriptive**	Argumentative	-3,5454	,94891	,002
		Comparison	-4,8612	,94891	,000
		Cause-Effect	-9,4875	,97409	,000
	Cause-Effect	Argumentative	5,9421	,93364	,000
		Descriptive	9,4875	,97409	,000
		Comparison	4,6263	,93364	,000

In terms of the productive use of words available among the second 1,000 words, both descriptive and cause-effect essays significantly differ from the remainder essay types.

Table 4. Descriptive Statistics for the Distribution of Basewords Included in First 1,000 Word Frequency Band across Datasets

	Essay Types	Mean	SD
1K Basewords	Argumentative	87,9716	2,40773
	Descriptive	85,7856	2,36661
	Comparison-Contrast	83,1437	3,26442
	Cause-Effect	83,6835	3,29847
	Total	85,1603	3,43307

The ratio of basewords to the total number of tokens among the first 1,000 words is almost 88% in argumentative essays which contain the largest number of words that do not have any affixes.

Table 5. MANOVA Test Results for Basewords Included in First 1,000 Word Frequency Band

	Essay Types	Essay Types	Mean Difference	Std. Error	Sig.
1K Basewords	**Argumentative**	Comparison	4,8279	,93316	,000
		Cause-Effect	4,2880	,96021	,000

The findings indicate that the number of productive use of basewords among the first 1,000 word frequency band in argumentative essays significantly differs from both comparison-contrast essays and cause-effect essays.

Table 6. Descriptive Statistics for the Distribution of Inflections Included in First 1,000 Word Frequency Band across Datasets

	Essay Types	Mean	SD
1K Inflections	Argumentative	8,4747	1,64163
	Descriptive	10,0587	1,82467
	Comparison-Contrast	11,1021	3,16981
	Cause-Effect	10,8888	2,47644
	Total	10,1128	2,55593

When it comes to the descriptive results about words with inflectional suffixes which are among the first 1,000 words, it is apparent that the type of essay, in which the number of words with inflectional suffixes is the least, is argumentative ones. Comparison-contrast essays containing the least number of basewords are the ones with the highest number of words with inflectional suffixes.

Table 7. MANOVA Test Results for Inflected Words Included in First 1,000 Word Frequency Band

	Essay Types	Essay Types	Mean Difference	Std. Error	Sig.
1K Inflections	**Argumentative**	Comparison	-2,6274	,77005	,007
		Cause-Effect	-2,4141	,79238	,020

The ratio of words with inflectional suffixes in the first 1,000 word frequency band to the total number of words is the least in argumentative essays. The use of affixes with the words in the first 1,000 word frequency band in argumentative essays differs significantly from the use of inflections with the words in the same word frequency band in comparison-contrast and cause-effect essays.

Tablo 8. Descriptive Statistics for the Distribution of Derivations Included in First 1,000 Word Frequency Band across Datasets

	Essay Types	**Mean**	**SD**
1K Derivations	Argumentative	3,5532	1,63713
	Descriptive	4,1550	1,65359
	Comparison-Contrast	5,7542	1,95370
	Cause-Effect	5,4247	1,74658
	Total	4,7259	1,95123

With respect to the use of derivational affixes with words of first 1,000 word frequency band, it is seen that comparison-contrast and cause-effect essays are the ones in which the words derived with affixes are used the most. However, the least frequently used words with derivational affixes in this word frequency band are available in argumentative essays.

Table 9. MANOVA Test Results for Derived Words Included in First 1,000 Word Frequency Band

	Essay Types	Essay Types	Mean Difference	Std. Error	Sig.
1K Derivations	**Comparison**	Argumentative	2,2011	,56996	,002
	Cause-Effect	Argumentative	1,8715	,58649	,013

The use of derived words in the first 1,000 word frequency band in both comparison-contrast essays and cause-effect essays differs significantly from those in argumentative essays.

Table 10. Descriptive Statistics for the Distribution of Basewords Included in Second 1,000 Word Frequency Band across Datasets

	Essay Types	Mean	SD
2K Basewords	Argumentative	40,1942	16,72560
	Descriptive	50,5731	13,41843
	Comparison-Contrast	39,6353	12,19734
	Cause-Effect	56,1629	7,99498
	Total	46,2070	14,60648

Table 11. MANOVA Test Results for Basewords Included in Second 1,000 Word Frequency Band

Independent Variable	Essay Types	Essay Types	Mean Difference	Std. Error	Sig.
2K Basewords	**Cause-Effect**	Argumentative	15,9687	4,29924	,006
		Comparison	16,5277	3,40444	,000

As for the productive use of basewords among the second 1,000 word frequency band, cause-effect essays contain the largest number of basewords and statistically differ from both argumentative essays and comparison-contrast essays.

Table 12. Descriptive Statistics for the Distribution of Inflected Words Included in Second 1,000 Word Frequency Band across Datasets

	Essay Types	Mean	SD
2K Inflections	Argumentative	29,8121	8,36169
	Descriptive	26,5125	9,72419
	Comparison-Contrast	41,7800	10,02535
	Cause-Effect	26,2988	7,44688
	Total	31,4300	10,88440

Table 13. MANOVA Test Results for Inflected Words Included in Second 1,000 Word Frequency Band

	Essay Types	Essay Types	Mean Difference	Std. Error	Sig.
2K Inflections	**Comparison**	Argumentative	11,9679	2,90552	,001
		Descriptive	15,2675	3,03866	,000
		Cause-Effect	15,4812	2,98975	,000

With regard to the inflected words which belong to the second 1,000 word frequency band, the essays containing the words with the highest number of inflectional suffixes are comparison-contrast essays which statistically differ from the rest of the essay types.

Table 14. Descriptive Statistics for the Distribution of Derived Words Included in Second 1,000 Word Frequency Band across Datasets

	Essay Types	Mean	SD
2K Derivations	Argumentative	29,9942	12,83763
	Descriptive	22,9156	11,40389
	Comparison-Contrast	18,5837	10,29215
	Cause-Effect	17,5376	5,03925
	Total	22,3630	11,34448

Table 15. MANOVA Test Results for Derived Words Included in Second 1,000 Word Frequency Band

	Essay Types	Essay Types	Mean Difference	Std. Error	Sig.
2K Derivations	**Argumentative**	Comparison	11,4105	3,37036	,007
		Cause-Effect	12,4566	3,46808	,004

In terms of the productive use of derivational affixes with words of second 1,000 word frequency band, argumentative essays are those that contain the largest number of derived words. MANOVA test results indicate that argumentative essays significantly differ from both comparison-contrast and cause-effect essays in terms of the use of derived forms of second 1,000 word frequency band.

Table 16. Distribution of Affixes in Argumentative Essays

				Affix Levels	Affix %	Text %
By Class	Basewords	80,03%	**By Bauer and Nation's Framework (1993)**	Level 2	62,88	11,97
				Level 3	11,78	2,25
	Inflections	11,97%		Level 4	9,32	1,78
	Derivations	7,06%		Level 5	7,28	1,39
				Level 6	8,53	1,63

When the dataset composed of argumentative essays is examined in general with regard to the affix profile of argumentative essays by class, it is revealed that 80,03% of the tokens in total are composed of basewords. However, inflected words

make up 11,97% of the data, while derivative words make up just 7,06% of the data. In addition, as for the affix profile of argumentative essays according to Bauer and Nation's framework (1993), regular inflections (Level 2) form 62,88% of the total number of affixes and 11,97% of the entire dataset. On the other hand, 2,25% of the total dataset and 11,78% of all affixes are generated by regular derivations (Level 3). Furthermore, 9,32% of the total number of affixes in the dataset is Level 4 derivations, while the ratio of these affixes in the total dataset is 1,78%. Moreover, while Level 5 derivations constitute 1,39% of the total dataset and 7,28% of the total number of affixes; Level 6 derivations constitute 1,63% of the total dataset and 8,53% of all the affixes.

Table 17. Distribution of Affixes in Descriptive Essays

By Class	Basewords	83,31%	**By Bauer and Nation's Framework (1993)**	Affix Levels	Affix %	Text %
				Level 2	66,57	11,15
				Level 3	10,65	1,79
	Inflections	11,15%		Level 4	9,99	1,68
	Derivations	5,59%		Level 5	10,35	1,74
				Level 6	2,20	0,38

The analysis of affix profile by class within the descriptive essays in general indicates that 83,31% of the total number of tokens are basewords which have neither prefixes nor suffixes. Additionally, 11,15% of the total number of tokens are words with inflectional suffixes. However, derived forms constitute just 5,59% of the total dataset. On the other hand, the affix profile of descriptive essays based on Bauer and Nation's framework (1993) shows that while regular inflections (Level 2) form 11,15% of the total number of words, they constitute 66,57% of the total number of affixes used. Additionally, while the ratio of regular derivations to the total number of affixes is 10,65%, the ratio of regular derivations to the total number of tokens in the dataset is 1,79%. The production rates of Level 4 and Level 5 affixes are very close to each other. While the ratio of Level 4 derivations to the total dataset is 1,68%, the ratio to the total number of affixes is 9.99%. These values are 1,74% and 10,35% for Level 5 derivations, respectively. Lastly, Level 6 derivations have little cumulative effect both on the total dataset (0,38%) and within the affixes used (2,20%).

Table 18. Distribution of Affixes in Comparison-Contrast Essays

By Class	Basewords	75,34%	**By Bauer and Nation's Framework**	Affix Levels	Affix %	Text %
				Level 2	64,30	14,88
				Level 3	9,42	2,19
	Inflections	14,88%		Level 4	7,31	1,70
	Derivations	8,25%		Level 5	10,49	2,44
				Level 6	8,30	1,93

The affix profile of comparison-contrast essays by class indicates that 75,34% of the total number of tokens consist of words that have neither inflectional nor derivational affixes. On the other hand, while the ratio of inflected words in the total dataset is 14,88%, the ratio of derivations in the total dataset is 8,25%. Additionally, the distribution of affixes based on Bauer and Nation's framework (1993) shows that although inflected words are 14,88% of the total number of tokens, 64,30% of the total number of affixes in the dataset are regular inflections. As for derivations, 2,19% of the total number of words in the dataset were created using regular derivative affixes (Level 3). However, the ratio of these regular derivative affixes to the total number of prefixes and suffixes used in the dataset is 9,42%. Also, derivations created by means of Level 4 affixes form only 1,70% of the total number of tokens although their ratio to the total number of affixes in the dataset is 7,31%. In addition, it seems that Level 5 derivative affixes are the ones used to derive the largest number of words in comparison-contrast essays in which 2,44% of the total number of words were produced with Level 5 suffixes which form 10,49% of the total number of affixes in the dataset. Finally, Level 6 derivative affixes make up 1,93% of the words in the dataset and 8,30% of the total prefixes and suffixes.

Table 19. Distribution of Affixes in Cause-Effect Essays

By Class	Basewords	76,95%	**By Bauer and Nation's Framework**	Affix Levels	Affix %	Text %
				Level 2	60,73	13,66
				Level 3	12,21	2,75
	Inflections	13,66%		Level 4	7,47	1,69
	Derivations	8,83%		Level 5	12,25	2,76
				Level 6	7,16	1,62

The affix profile analysis of the dataset which is formed of cause-effect essays by class indicates that 76,95% of the total number of tokens are basewords. However, while the ratio of words with inflectional suffixes to the total number of words is 13,66%, the ratio of words with derivational prefixes and suffixes to the whole dataset is 8,83%. On the other hand, the affix profile of cause-effect essays according

to Bauer and Nation's framework (1993) reveals that 60,73% of the total number of affixes in the dataset are regular inflectional suffixes. On the other hand, the ratio of words derived with Level 3 affixes, called regular derivational affixes, to the total number of tokens and the number of affixes used in the dataset is 2.75% and 12.21%, respectively. Moreover, the number of Level 4 derivational affixes is less than the Level 3 derivational affixes both in the total dataset (1,69%) and among the affixes (7,47%). Additionally, in this dataset of cause-effect essays, both the number of words derived from Level 5 affixes (2,76%) and the ratio of Level 5 affixes to the other prefixes and suffixes are the highest (12,25%). Lastly, Level 6 derivational affixes are the least in number both among the derived words (1,62%) and among the other affixes used in the dataset of cause-effect essays (7,16%).

Table 20. Distribution of the Most Frequent Affixes across Essay Types

Argumentative Essays		Descriptive Essays		Comparison-Contrast Essays		Cause-Effect Essays	
Derivative Affix	Percentage %	Derivative Affix	Percentage %	Derivative Affix	Percentage %	Derivative Affix	Percentage %
-ly	12,97	-ful	19,21	-ly	11,56	-ly	15,76
dis-	10,01	-ly	11,38	-al	9,15	-ion	13,73
-able	8,74	-y	9,43	-ion	8,67	dis-	6,98
-ion	8,74	-hood	9,07	-ic	7,10	-al	6,86
-al	6,34	-ion	4,98	-ent	5,18	-ive	6,64
in-	5,07			-er	4,21	-ation	4,72
				-ation	3,85		
				-ive	3,61		
Total	51,87	Total	54,07	Total	53,33	Total	54,69

Table 20 shows the most commonly used affixes, which are available in Bauer and Nation's framework (1993), in each essay type. The affixes mentioned in Table 20 constitute roughly 50% of the total number of affixes used in each dataset. When the ratio of the most frequently used affixes to the total number of affixes in each dataset exceeds 50%, the affixes that are used relatively less are not included in Table 20. It seems that suffixes contribute more to EFL learners' productive vocabulary than prefixes. Among the most frequently used affixes, which are reported in Table 20, constituting roughly 50% of the number of affixes used in each dataset, only dis- and in- are prefixes. On the other hand, the suffix -ly is the most frequently used affix in all essay types except the descriptive essays. In addition, according to Table 20, the suffixes -ly, -ion, -al and the prefix dis- are the affixes that contribute the most to the productive vocabulary of EFL learners.

DISCUSSION

It seems that the dataset, in which the first 2,000 words are used most frequently, is the one consisting of descriptive essays. On the other hand, the dataset in which the first 2,000 words are used the least is the one that includes comparison-contrast essays, implying that this dataset contains less frequent and more demanding words.

Productive use of the first and second 1,000 words differs significantly among the datasets. While the dataset consisting of descriptive essays differs from the other 3 datasets in terms of the productive use of the first 1,000 words, the dataset that is formed of cause-effect essays significantly differentiates from the rest of the datasets with regard to the productive use of the words included in the second 1,000 word frequency band. This finding suggests that the essay type has a significant effect on the productive use of the most frequent first and second 1,000 words. Interestingly, on the one hand, the productive use of the first 1,000 words most frequently in descriptive essays, which are written in a more informal manner compared to other essay types, draws attention; on the other hand, the scarcity of the first 1,000 words in cause-effect essays, in which a formal attitude is dominant, draws attention. Similarly, as for the productive use of words included in the second 1,000 word frequency band, similar conclusions can be drawn. In the descriptive essays, which are written with an informal attitude, less frequent words are used less frequently; however, in cause and effect essays, where the formal attitude is dominant, the use of less frequent words increases as the word frequency decreases. Therefore, it is possible to infer that the level of formality has an effect on the selection of words.

The findings also suggest that as the frequency band in which the words are included decreases, the number of basewords that do not have any affixes also decreases. The number of basewords with no affixes included in the second 1,000 words (M: 46,2070) corresponds to almost half of the number of basewords without any affixes included in the first 1,000 words (M: 85,1603). Accordingly, as the frequency band that the words are included in decreases, the number of words with inflectional and derivational affixes dramatically increases. The number of words with inflectional affixes from the words included in the second 1,000 words is 3 times more than the number of words included in the first 1,000 words with inflectional affixes. Similarly, the number of words with a derivational affix among less frequent words is almost five times the number of words with a derivational affix among the first 1,000 words. Therefore, it can be concluded that EFL learners tend to derive more derivatives included in the less frequent word bands by using derivational affixes compared to inflected words.

It is also possible to mention that the essay type has a partial effect on the use of words with and without affixes. As for the words included in the first 1,000 word frequency band, argumentative essays differ from both comparison-contrast essays and cause-effect essays in terms of the use of basewords, inflected words, and derived forms. Also, argumentative essays differ from abovementioned two essay types in terms of the use of derivatives included in the second 1,000 words. This finding is in line with Laufer and Cobb (2019) who come to the conclusion that the number of affixed words in different text types differs from each other. As the findings of the present study indicate, different types of essays contain different numbers of affixed words.

On the other hand, inflectional suffixes called Level 2 appear as the most frequently used suffixes in each dataset. More than 60% of the total affixes used in each dataset are inflectional suffixes. As for derivational affixes, Level 3 affixes, which are the most frequent and regular derivational affixes, are among the most frequently used affixes in each dataset along with Level 5 affixes which are regular but infrequent. Interestingly, even if they are infrequent affixes, Level 5 affixes are the ones which are the most frequently used affixes to derive words in comparison-contrast and cause-effect essays. However, in argumentative and descriptive essays, the most frequent and regular derivational affixes (Level 3) are mostly used to make derivatives. Also, the affixes which are frequent and orthographically regular (Level 4) and the ones which are frequent but irregular (Level 6) are among the less frequently used affixes in each dataset.

Morphology of a lexical item, more specifically inflectional complexity and derivational complexity are regarded to be among factors affecting word learnability in Laufer (1997). The findings of the present study suggest that inflectional affixes do not create as much learning burden as derivational affixes do, which also supports Schmitt and Zimmerman's (2002) suggestion that derivations and inflections do not cause the same amount of learning burden. There could be many reasons for this but the limited number of inflectional suffixes and teaching them as a part of grammar can be cited as one of the leading facilitating factors. Kim (2013), for instance, emphasizes the importance of explicit instruction of affixation on the improvement of EFL learners' vocabulary. In addition, in the present study, essays written by EFL learners with C1 language proficiency level were used as data. Having advanced language skills can be another reason to be proficient in the use of inflectional suffixes. It is also possible to conclude that the most frequently used members in word families are those with inflectional suffixes.

The findings of the present study towards derivational affixes (Level 3-6) suggest that essay writers with C1 level English proficiency are not more likely to use a particular group of affixes. Contrary to the expectation, Level 3 affixes which are the most frequent and regular derivational affixes are not the ones which are most

frequently used in each essay types, implying that EFL learners with C1 level of English proficiency are skilled in deriving words by using derivational affixes with different difficulty levels. For instance, irregular affixes (Level 6) are used more than orthographically regular affixes (Level 4) to derive words in comparison-contrast essays. Similarly, infrequent derivational affixes (Level 5) are the most frequently used affixes, more than even the most frequent and regular derivational affixes (Level 3), to derive words in cause-effect essays and comparison-contrast essays. It seems that these findings do not support Laufer (1997) underlining that lack of regularity and deceptive transparency of meaning create problems for learners because derivations are not derived with fixed rules such as inflected words and it is not always easy to guess the meaning that prefixes/suffixes add to the word. The findings of the present study also contradicts Leontjev's data (2016) which found that an increase in the affix level mentioned by Bauer and Nation (1993) causes a decrease in the recognition level of affixes. Maybe the findings of the present study can be explained with Iwaizumi and Webb's (2021) findings concluding that as the language proficiency increases, productive derivative vocabulary knowledge also increases. In Iwaizumi and Webb (2021), L1 speakers of English produce much more derivatives than ESL graduates and EFL undergraduates do. On the other hand, the number of derived words produced by ESL graduates outnumbers derived words produced by EFL undergraduates, suggesting that more proficient language users are better at producing derivatives. Similarly, Leontjev et al. (2016) emphasize the relationship between language proficiency and word derivational knowledge, clearly underlining that some aspects of word derivational knowledge escalate rapidly when the learners go above A2 or B1 level of English proficiency.

It is also clear that the words derived by means of suffixes outnumber the ones formed through prefixes. The number of words derived with prefixes is low due to the scarcity of prefixes used in the dataset. The adverb-making suffix -ly, which is mentioned to be one of the most basic level affixes in Laufer and Cobb (2020) who also find that the suffix -ly is among the affixes which form the 50% of the derivations in their data, appears to be the most frequently used affix in almost all types of essays. In addition, the findings of the present study partly support Mochizuki and Aizawa (2000) who claim that well-known affixes are acquired earlier. Both the present study and Mochizuki and Aizawa (2000) find that the suffixes -ation and -ful are among the well-known affixes.

CONCLUSION

The present study was conducted to reveal the role of affixation on productive vocabulary knowledge in different essay types. More specifically, it explored the distribution of affixed words within the most frequent first and second 1,000 words. It also examined how affixes forming lexical items disperse among the types menti-

oned by Bauer and Nation (1993). The most frequently used derivative affixes were further examined. The findings indicate that the essay type has a significant role on the productive use of the most frequent first and second 1,000 words. Additionally, as the words become less frequent, the number of inflectional and derivational affixes attached to these words dramatically increases. On the other hand, the use of derivative suffixes is not consistent with Bauer and Nation's framework (1993). Also, the effect of prefixes on the affixation process is quite limited.

These findings might have some pedagogical implications for language teachers and curriculum designers. The current study demonstrates the importance of affixation in low-frequency words, leading to conclude that more emphasis should be placed on teaching affixes to teach less frequent words. Explicit efforts to teach the meanings of affixes and to increase learners' awareness toward the morphological structures of words can be beneficial in teaching less frequently used words. In addition, the findings indicate that the ranking in Bauer and Nation's framework (1993) does not accurately reflect the order of difficulty for learners with C1 level of foreign language proficiency. Therefore, it may not make much sense to categorize affixes based on difficulty and to follow any order in their teaching. It should be noted that this inference may not be valid for learners with different proficiency levels such as beginner and intermediate. Namely, the ranking in Bauer and Nation's framework (1993) might really reflect the order of difficulty for less proficient English learners. Also, the findings point to the weakness of the role of prefixes in word derivation. Putting more emphasis on prefixes while teaching affixes can provide more fruitful results in teaching vocabulary. Morphological analyses to enable learners to distinguish prefixes might play an important role in improving their vocabulary knowledge.

Conflict of Interest

The author has no conflict of interest to declare.

Author Contributions

The author accepts full responsibility for the work.

REFERENCES

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography,* 6, 253-279.

Cobb, T. Compleat Web VP v.2.5 [computer program]. Accessed 1 Jun 2022 at https://www.lextutor.ca/vp/comp/

Cobb, T. Morpholex Affix Profiler v.3.5.1 [computer program]. Accessed 1 Jun 2022 at https://www.lextutor.ca/cgi-bin/morpho/lex/

Danilović, J., Savić, J., & Dimitrijević, M. (2013). Affix acquisition order in Serbian EFL learners. *Romanian Journal of English Studies,* 10(1) 77-88. https://doi.org/10.2478/rjes-2013-0006

Hayashi, Y., & Murphy, V. (2011). An investigation of morphological awareness in Japanese learners of English. *Language Learning Journal,* 39(1),105-120.

Kim, C. (2013). Vocabulary acquisition with affixation: Learning English words based on prefixes and suffixes. *Second Language Studies,* 31(2), 43–80.

Iwaizumi, E., & Webb, S. (2021). To what extent does productive derivational knowledge of adult L1 speakers and L2 learners at two educational levels differ? *TESOL Journal*, 12(4), e640. https://doi.org/10.1002/tesj.640

Laufer, B. (1997). What's in a word that makes it hard or easy? Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140–155). Cambridge: Cambridge University Press.

Laufer, B., & Cobb, T. (2020). How much knowledge of derived words is needed for reading? *Applied Linguistics,* 41(6), 971-998. https://doi.org/10.1093/applin/amz051

Leontjev, D. (2016). L2 English derivational knowledge: Which affixes are learners more likely to recognize? *Studies in Second Language Learning and Teaching,* 6(2), 225-248.

Leontjev, D., Huhta, A., & Mantyla, K. (2016). Word derivational knowledge and writing proficiency: How do they link? *System,* 59, 73-89. https://dx.doi.org/10.1016/j.system.2016.03.013

Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System,* 28(2), 291–304. https://doi.org/10.1016/S0346-251X(00)00013-0

Nation, I. (2013). Learning vocabulary in another language (Cambridge Applied Linguistics. Cambridge University Press. https://doi:10.1017/CBO9781139858656.003

Nation, P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 15-29). Routledge.

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36(2), 145-171. https://doi.org/10.2307/3588328

Tamura, T., & Shirahata, T. (2016). Prefix difficulty order among Japanese university learners of English. *Studies in Subject Development*, 4, 47–56.

Tamura, T., & Shirahata, T. (2017). Knowledge of English prefixes among Japanese adult learners of English. JACET Journal, 61, 69-87.