Investigating the Use of Nominalization: A Comparative Study of Pakistani Learners' and Native English Speakers' Argumentative Essays

Aisha Aslam[1], Dr Shafaq Fayyaz [2]

[1] M.Phil in Applied Linguistics Scholar, University of Lahore, Pakistan

[2] Assistant Professor, Department of English Language and Literature, University of Lahore, Pakistan

ABSTRACT

This research aimed at investigating the frequency of occurrence of nominalization in the argumentative essays written by Pakistani undergraduates and English native speakers. The corpus based analysis using log likelihood (LL) and Bayesian Information Criterion (BIC) showed that the frequency differences of nominalization in argumentative essays written by Pakistani students (PAK) and the argumentative essays written by English native speakers (ENS) are statistically significant. The study has revealed that although there is a statistically significant difference between the frequencies of nominalizations in PAK and ENS, this difference is not evenly distributed across the two corpora. In other words, out of the four different types of nominalizations ending in -tion, -ment, -ity and -ness, nominalizations ending in suffix -tion were 'overused' in PAK whereas nominalizations ending in suffix -ment were 'overused' in ENS. This clearly depicts the lexical preferences shown by Pakistani undergraduates and English native speakers. Significant contributions of this study are the benchmarking it provides for more diverse and in depth studies on nominalizations in the context of Pakistani Academic English, and offering informed insights for pedagogical implications.

Keywords

nominalization; English for Academic Purposes; Pakistani Academic English; Argumentative Essay; comparative corpus analysis; professional discourse

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

Introduction

As everyday language is distinct from the formal academic text, many learners face challenges transforming it into formal expressions, but relying on features like nominalization and a more noun oriented text approach may prove to be helpful (Liardét & Black, 2020).

Numerous studies have been conducted on nominalizations used by non-native users of English. Multiple studies have been done on the use of nominalizations by non-native users of English ranging from Chinese, Korean, Iranian and Japanese learners of English (Baratta, 2010; Khamesian,2015; Yue et al., 2018; Jalilifar, Alipour, & Parsa, 2013). The western, Chinese or Iranian researchers have explored nominalization in various academic contexts; however, not much focus has been given to studying this significant feature of academic writing by Pakistani researchers. Over the past few years, multidimensional analysis has gained popularity amongst Pakistani researchers (Abdulaziz, Mahmood, & Azhar, 2016; Ali & Shehzad, 2019; Mahmood, & Hussain, 2016; Aziz ,Pathan &Ali, 2017); however, there is not much that has been done to explore nominalizations in the writing of Pakistani learners of English.

Motivated by this lack of focused research on nominalizations in Pakistani Academic English context, this research aims at getting an informed insight into the frequency of use of nominalizations in Pakistani undergraduates' argumentative essays (henceforth referred as PAK) and to compare the results with a corresponding corpus of native English speakers (henceforth referred as ENS). This insight has pedagogical implications for teachers and it provides a vantage point for the researchers to explore nominalization in the context of Pakistani English.

The research has following objectives:

a. Identify and count the frequency of occurrence of nominalizations in PAK and compare it with ENS.
b. Report any statistically significant frequency differences between PAK and ENS.

It is hypothesized that;

There is a significant difference between the frequencies of nominalization used in PAK and ENS.

Nominalization in Academic English

Hinkel (2002) analyzed a large pool of L2 texts and compared them to native speakers' texts. She investigated nominalizations labelled as 'abstract generic nouns', included all nouns that end in suffixes '-s/tion, -ment, -ness, -ure, -ity' as nominalized forms, and listed them under the category of linguistic features along with the other 15 features that she analyzed. The results showed that there were no significant differences in the frequency of use of nominalization in the native speakers' and non-native speakers' essays with the exception of Vietnamese and Indonesian learners who used nominalization at a very low rate. Hinkel (2013) updated her definition of nominalizations as she excluded the suffix '-ure' and used the suffixes that mark nominalization listed by Biber (1988) in his factor analysis. In an interesting study conducted on six undergraduates' academic writing development throughout each year of their undergraduate studies, Baratta (2010) analyzed if effective use of nominalization plays an important role in the development of their academic writing. He concluded that at least for the group of students and the disciplines and degree programs he studied, it did not play a very important role in academic writing proficiency until the final year of their degree program. Biber and Gray (2013) conducted a corpus analysis on science research articles written in the twentieth century. They claimed that nominalizations are one of the most unique features of the modern academic writing (Biber & Gray, 2013). Their study showed that the increase in nominalization use does not correspond to the decreased use of verbs throughout every genre. Their data comprised of four genres; newspapers, academic prose, drama and fiction. Within the academic prose, they focused more on the scientific research articles. Hence, as for essay writing, we do not find any information in this study. Studies on analyzing nominalizations in the books of applied linguistics and biology showed no significant difference in the use of nominalization in the books of the two disciplines (Jalilifar et al., 2014, 2017). Some studies have explored nominalization in different parts of research articles and masters and doctoral dissertations (See Yue et al., 2018; Yue and Zhang, 2019; Sarani and Talati-Baghsiahi, 2015).

Two studies that give insights into nominalizations in argumentative essays are done by Kim and Nam (2019) and Yoon (2018) where the latter is more in-depth in terms of exploring nominalizations.

Some recent studies in Pakistani Academic Writing context have made some references to the role of estimated frequencies of nominalizations and most of these studies suggest that the writing of Pakistani students is marked by a comparatively higher use of nominalizations (Abdulaziz et al., 2016; Azher et al., 2017); Azher et al., 2019; Tabassum et al., 2019). As there seems to be no published studies found exclusively exploring nominalizations in the academic writing of Pakistani learners of English language, this study aims to bridge that gap.

Conceptual Framework

As Figure 1 illustrates clearly, nominalization in this study is defined after the basic definition adopted by Biber (1988), and the comparative quantitative study of this lexico-grammatical feature in the argumentative essays of Pakistani undergraduates and English native speakers, yields two major outcomes; similarities and differences between the frequency of occurrence of nominalization in the two sets of corpora under investigation. Although both the similarities as well as the differences in the frequency of occurrence of nominalization maybe significant, the differences offer more insightful patterns of use by the Pakistani students. A careful study of the differences between the ways nominalizations are used by Pakistani and English native speakers offers two principal benefits. Firstly, the informed insight allows English for Academic Purposes teachers to plan more effective ways to teaching nominalizations by focusing more on the forms of nominalizations that are either underused or overused by Pakistani undergraduates; secondly, the insight also helps benchmarking any future corpus based studies on nominalizations in Pakistani Academic English writing by offering quantitative data that could be compared with a similar corpus to get further insights into Pakistani students' use of nominalization in academic English.

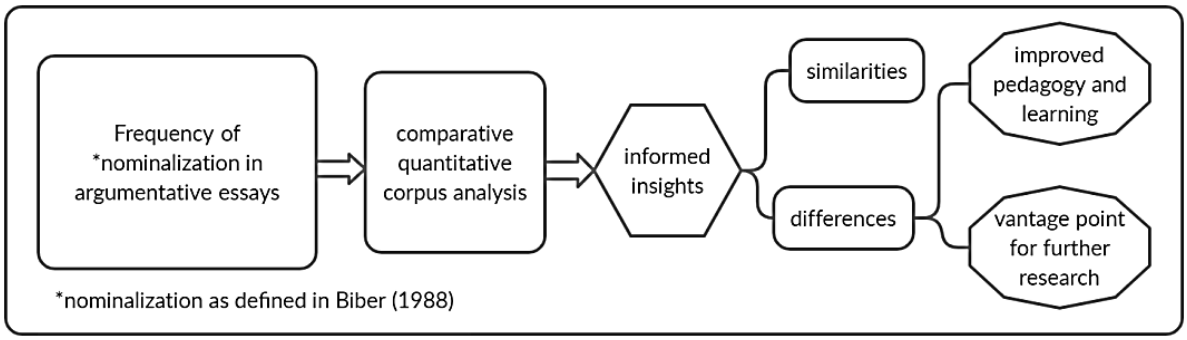


Figure 1. Conceptual framework of the study

Methodology

The current research is a corpus based comparative study and is mainly quantitative in its design.

Data Collection: The Sub-corpora Selection and Modification

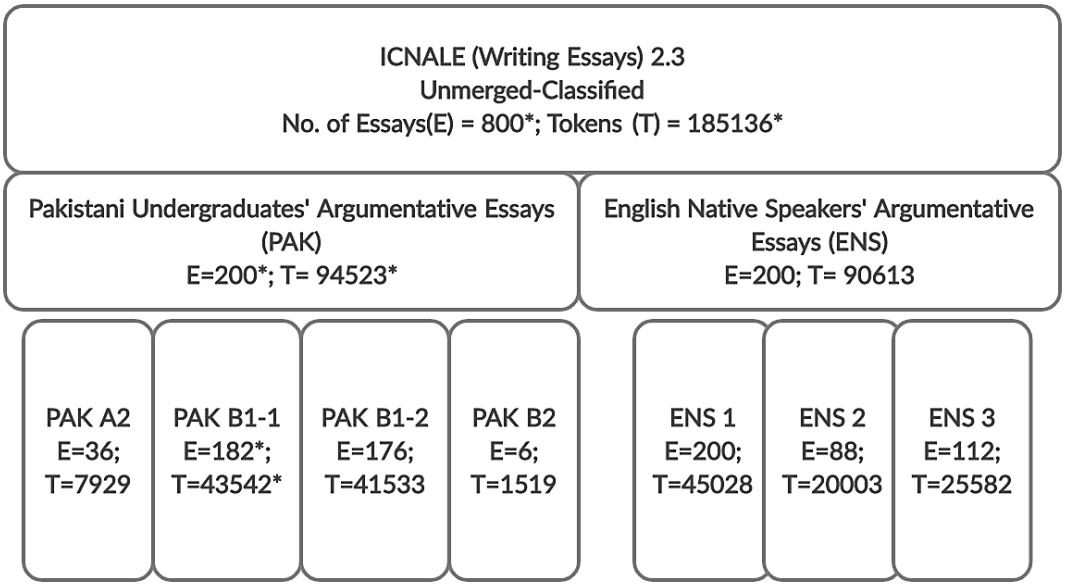


Figure 2. Sub-corpora Extracted from ICNALE.

The 'Written Essays' (c 1.3 million tokens) module of the larger corpus of International Corpus Network of Asian Learners of English (ICNALE) is the data source for this research. From this module of ICNALE, Pakistani learners' argumentative essays and native English speakers' argumentative essays are extracted and for the sake of this research these two sub-corpora are labelled as PAK and ENS respectively. Figure 2 illustrates the details of the total number of essays and token count in PAK and ENS.

Note: The data with (*) depict the original numbers before the elimination of a file in PAK

While compiling ICNALE, essay prompts were not the only variables that were kept constant, but also the other variables such as learners' proficiency levels, time to write the essays, the length of the essays were also strictly controlled (Ishikawa, 2018). For the purpose of this research, different proficiency levels were not a variable under consideration, hence the entire PAK data that comprise of essays from learners at various proficiency levels are treated as one organic whole. There were two essay prompts given to the students to write two argumentative essays.

(a) It is important for college students to have a part-time job (Referred to as PTJ).
(b) Smoking should be completely banned at all the restaurants in the country (Referred to as SMK).

Two files in PK-B1-1(Figure 2) had duplicated content, hence one of them (W_PAK_PTJ0_131_B1_1) was removed and the token count and the number of essays were adjusted. The final token count for the entire PAK turned out to be 94,312 (cf 94523*, Figure 1). The accumulative token count for the research also became 184,925 tokens (cf 185,136*, Figure 2) and the total essay files dropped to 799 instead of 800 original files that were extracted from ICNALE.

The Exclusion Criteria

Researchers in the past have defined and studied nominalization from various perspectives, hence it is not as straightforward to retrieve this lexico-grammatical feature in any given corpus. In order to make the research more manageable, it is imperative that standardization criteria are devised for the identification of nominalization.

This research followed the basic definition of nominalization given by Biber(1988). According to Biber(1988, p.227) " any noun ending in -tion, -ment, -ness, or -ity plus the plural forms" count for nominalizations. However, problems arise when the data are tagged by software that is not solely written to identify nominalization from a particular perspective. For instance, when searching for the suffix -ment in the current research data, AntConc (3.5.7.0) returned a range of items that are not nominalizations. Words like 'mentioned', 'mentioning', 'government', 'compartment' etc. were all tagged under -ment suffix category. Therefore, well planned exclusion criteria were needed. For this research, the following exclusion criteria were devised (See Appendix A for detailed list of words excluded from the corpus based on these criteria):

a. If a word is not derived from a verb or adjective (adapted from Biber et al (1999)), it is not considered as nominalization. For example, words such as 'question', 'caution', 'nation', and 'transition'.
b. Any concrete nouns ending in -tion, -ment, -ness, or -ity , such as, 'partition', 'instrument', 'section', 'harness (n)' etc. are also not considered as nominalizations.
c.

Procedures, Tools and Description

The following statistics were utilized in this research:

(a) Raw frequencies, normalized frequencies, percentages
(b) Statistical Significance and Effect Size Test

Raw frequencies along with the percentages offer an intuitive picture of the usage of a certain linguistic element used in a corpus with respect to the other related corpus (Nausa, 2019).

Normalized frequencies are used when the size of the corpora are not exactly the same. Normalized frequencies depict an item's overuse or its underuse in a corpus or sub-corpus in relation to another. For example, the sample essays used in this research strictly follow the word limit; each essay is between 200- 300 words and the students were given 20-40 minutes to write them. However, some variation between the word counts is inevitable. Therefore, the frequencies were normalized to 1000. The following formula suggested by Biber (1988) was used:

(Actual frequency count ÷ total words in text) × 1000

These normalized frequencies give a better and more efficient overall relative frequency count of a linguistic item as compared to the raw frequencies and percentages. Normalized frequency counts are more efficient way of measuring the occurrence of an item when comparing its use across different corpora; however, they are not sufficient to test the hypothesis. To test a hypothesis, statistical analysis is needed. A general method utilized in corpus linguistics is to use inferential statistics tests, such as statistical significance and effect size tests.

For this research, log likelihood (LL) was calculated to measure statistical significance of any differences found in the use of nominalizations used in PAK in relation to ENS. LL measurement provides the evidence for statistical significance but it does not indicate how much (or the magnitude) of a difference that really is. For the purpose of calculating the magnitude of the difference Bayesian Information Criterion (BIC) was used as approximate values

because true Bayes Factors (BF) are not that straight forward to calculate (Nausa, 2019). According to Nausa (2019) another reason to use BIC values is that they test the possibility of a hypothesis based on data that is available and hence provide the degrees of evidence against null hypothesis (H_0). For log likelihood measurement and Bayesian Information Criterion calculation an online calculator was used (http://ucrel.lancs.ac.uk/llwizard.html).

Software Used

LL and BIC values cannot be calculated without the frequency count of the linguistic items under investigation. Therefore, in order to calculate the raw frequency of nominalizations used in PAK and ENS, AntConc (3.5.7) was used on a Windows 10 computer. AntConc (3.5.7) was used to perform the following functions.

(a) Generate concordance list
(b) Generate word list

In order to calculate the frequencies of nominalizations ending in -tion, -ment, -ity, -ness and their plurals, the 'Regex' function in AntConc (3.5.7) was used. The exclusion criteria were applied to the results of 'Regex' search and the final raw frequency count was calculated. To record and calculate the frequencies word lists, based on AntConc(3.5.7) setting 'sort by word end' were also generated. These word lists were later imported to Microsoft Excel for further calculations, normalization, and for creating relevant graphs and charts for data representation.

Analysis and Findings

The sub-corpora; PAK and ENS were analyzed quantitatively and the following results were obtained.

Raw and Normalized Frequencies in PAK and ENS

Table 1 shows the raw frequencies, normalized frequencies along with the percentages of nominalization ending in the four suffixes (-tion, -ment, -ity, -ness).

Table 1

Raw Frequencies and Normalized Frequencies (Nf) of Nominalizations in PAK and ENS

	Tokens	NOM -tion	Nf	NOM -ment	Nf	NOM -ity	Nf	NOM -ness	Nf
PAK	1363	619 (45%)	454.15	341 (25%)	250.18	335 (25%)	245.78	68 (5%)	49.89
ENS	1124	405 (36%)	360.32	399 (36%)	354.98	272 (24%)	241.99	48 (4%)	42.70

Note. Total Tokens PAK= 94312, ENS= 90613;
PAK (Nf=14.45), ENS (Nf=12.40)

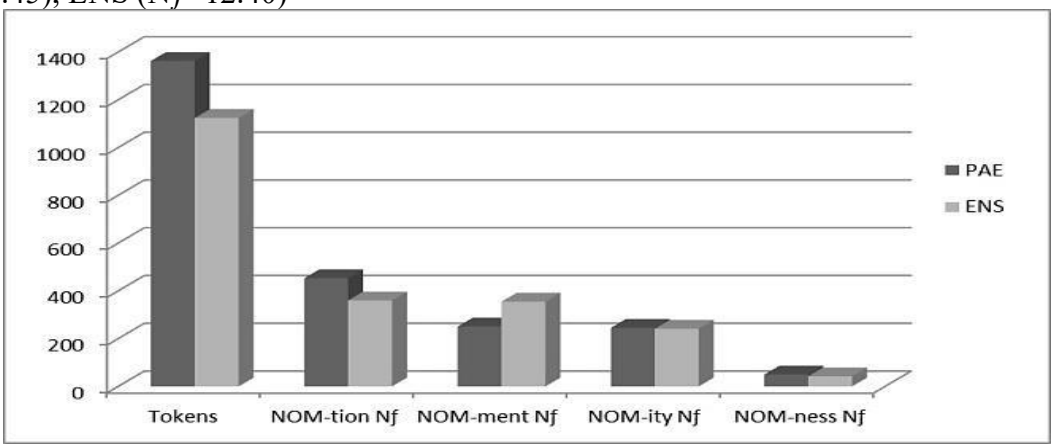


Figure 3. Tokens and Normalized Frequencies of Nominalizations in PAK and ENS

Both Table 1 and Figure 3 illustrate the differences between normalized frequencies of nominalizations across PAK and ENS .While the difference between overall token count (raw frequencies) clearly depicts a higher usage of nominalizations in PAK as compared to ENS, a more closer look shows that this 'overuse' is not evenly distributed across all four categories of nominalizations. In fact, nominalizations ending in -ment (Henceforth 'NOM-ment') are used more frequently in ENS. Moreover, the differences between nominalizations ending in suffixes -ity and –ness (Henceforth NOM-ity and NOM-ness respectively) do not seem to be significant where PAK has an almost negligibly higher normalized frequency of occurrence compared to ENS. The only category of nominalizations where PAK has a noticeably higher usage is NOM-tion (Henceforth 'NOM-tion'). Hence it can be concluded that this general analysis shows an overall picture of the raw and normalized frequencies across PAK and ENS and is valuable for an overall comparison of the sub-corpora, however, it is not sufficient enough to test the hypothesis.

Statistical Significance and Effect Size

The hypothesis for this research is as follows;

There is a significant difference between the frequencies of nominalizations used in PAK and ENS.

In order to test the hypothesis, inferential statistics tests are needed. Therefore, in this research, log likelihood (LL) and Bayesian Information Criterion (BIC) tests were used for statistical significance and effect size calculations respectively.

Table 2

Log Likelihood and Bayes Factor Tests for Nominalizations in PAK Relative to ENS

	Corpus Size	Raw*f* of NOM	log likelihood(LL)	+/-	Sig.	Bayesian Information Criterion (BIC)
PAK	94312	1363	14.44	+	***	2.31
ENS	90613	1124				

Note. "+" indicates overuse in PAK relative to ENS; Sig.= Significance (*=p<0.05, **=p<0.01, ***= p<0.001); BIC (2-6: Positive evidence against H_0, 6-10: Strong evidence against H_0)

Table 2 depicts the difference between the frequencies of nominalizations used in PAK and ENS is statistically significant with a positive LL value of 14.44 (p<0.001). The differences in frequencies depicted by the normalized frequencies and the percentages are also reinforced by this statistical analysis.

In addition to the frequency difference between the two sub-corpora being analyzed for the use of nominalizations, the effect size calculation using Bayesian Information Criterion (BIC) provides positive evidence to disapprove null hypothesis (H_0), and as a consequence the hypothesis that there is a significant difference between the use of nominalizations in PAK and ENS is therefore approved.

Moreover, in order to explore further patterns of frequency distribution in different categories of nominalizations based on the four suffixes, the statistical significance of each category in PAK was calculated with respect to ENS. The LL values gave the evidence for a category of nominalization being overused or underused in PAK with respect to ENS and that information was sufficient to decide if a particular category was more statistically significant than the others.

Table 3
Log likelihood of Different Nominalizations in PAK with Respect to ENS

	PAK	ENS	log likelihood(LL)	+/-	Sig.
Raw *f* of nominalizations	1363	1124	14.44	+	***
NOM-tion	619	405	13.30	+	***
NOM-ment	341	399	22.60	-	***
NOM-ity	335	272	0.04	+	-
NOM-ness	68	48	0.69	+	-

Note. "+" shows "overuse" of nominalization in PAK in relation with ENS, while "-" depicts "underuse" of nominalization in PAK in relation with ENS; Sig.= Significance (*=p<0.05, **=p<0.01, ***= p<0.001)

An overall comparison of Table 2 and Table 3 shows that although the difference of frequencies of nominalization used in PAK in relation to ENS is statistically significant, not every category of nominalizations analyzed in this research is statistically significant in terms of the differences of their frequency. Table 3 illustrates that fact in a more detailed manner. In fact, the only category of nominalization where the overuse of nominalization in PAK is statistically significant is NOM-tion. Interestingly, although the hypothesis that there is a significant difference between the use of nominalization in PAK and ENS is already accepted (Table 2), it can be clearly seen in Table 3 that two of the categories; NOM-ity and NOM-ness do not show statistically significant differences of frequency across both the sub-corpora. This is valuable evidence as it illustrates that not all forms of nominalizations are used the same way in PAK and ENS. (See Appendix C for the raw results of LL and BIC for both Tables 2 and 3).

Discussion

It was hypothesized that there is a significant difference between the frequencies of nominalizations in PAK and ENS. The findings of the research; BIC (2.31) and LL (+14.44) have proved the hypothesis to be true.

The answer to the first research question illustrated an overall difference between the frequencies of occurrence of nominalization in PAK (N*f*=14.45) and ENS (N*f*=12.40). As for the answer to the second research question, the findings of this research show that there is statistically significant difference between nominalizations used in argumentative essays written by Pakistani students and native English speakers'. This result contradicts Hinkel's (2002) research where she concluded that there were no significant differences in the frequency of use of nominalization in the native speakers and non-native speakers' essays with the exception of Vietnamese and Indonesian learners who used nominalization at a very low rate. Similar results are reported in an older research where no significant differences between the frequencies of nominalizations were reported in native speakers' and non-native speakers' academic essays (Carlson, 1988 as cited in Hinkel, 2002). These conclusions are reiterated in a study done by Kazemi (2015) on articles written in the field of English Language Teaching by Iranian and English native speakers. He concluded that there was no significant difference between the use of nominalizations by the Iranian writers and English native speakers in their articles (also see Yoon, 2018; Sarani & Talati-Baghsiahi, 2015).

On the other hand, a significant difference of frequencies of nominalization between the writings of native and non-native speakers was noted by Liu et al.(2014) .They concluded that the Chinese writers used more nominalizations as compared to the British writers; similar to Pakistani writers that used overall more nominalizations in their essays when compared with the native English speakers. A similar study done by Abdulaziz et al.(2016) investigated variations in learners' argumentative essays extracted from International Corpus Network of Asian Learners of English (ICNALE) and concluded that Pakistani learners' writing is

'heavily nominalized' and more informational compared to English native speakers' corresponding text. The present research results affirm Abdulaziz et al. (2016) in one way; there is a significant difference between the nominalizations used by Pakistani students and English native speakers essays. However, Pakistani students' essays are not 'heavily nominalized'. In fact, the quantitative analysis results show that it is only NOM-tion that is significantly higher in its frequency of usage compared to its use in English native speakers' essays (See Figure 3). Overall, PAK shows an overuse of nominalizations (See Table 1) compared with ENS, but ENS has overused NOM-ment when compared with PAK. Hence, this proves that Pakistani students' argumentative essays are not 'heavily nominalized', although the difference between the overall use of nominalizations between PAK and ENS is statistically significant. These results also contradict with the conclusions made by Abdulaziz et al. (2016) that Pakistani students' argumentative essays are loaded with nominalizations and that they are more information packed.

Moreover, the findings have shown that the nominalization ending in different suffixes do not have the same frequency of usage in both the sub-corpora analyzed. In other words, the results have shown that NOM-tion has the highest frequency in PAK and ENS. The highest frequency of NOM-tion in PAK and ENS is in accordance with Biber et al.(1999) where they concluded that suffix -tion is the most productive suffix used to make nouns in academic prose. It can be observed clearly that in terms of the raw frequencies, -tion is used most frequently as the suffix of nominalizations found in both PAK and ENS.

Interestingly, the nominalizations created using -ment are far greater than one would expect to see considering Biber et al. (1999). In both PAK and ENS, the most frequently occurring nominalization has the suffix -tion but it is followed by -ment nominalizations instead of -ity as shown by Biber et al. (1999). Additionally, -ment is the least commonly used noun forming suffix according to Biber et al.(1999) when compared with the other 3 (-tion, -ity , and -ness) noun making suffixes. Moreover, nominalizations ending in -ness are the least frequent in both PAK and ENS (only 5% and 4% respectively).

Conclusion

This research investigated the frequency of occurrence of nominalizations in a corpus of argumentative essays written by Pakistani undergraduates and compared those frequencies with a parallel corpus of English native speakers' argumentative essays. The corpus based analysis employing log likelihood (LL) and Bayesian Information Criterion (BIC) showed that the frequency differences of nominalizations in PAK and ENS are statistically significant. The study has revealed that although there is a statistically significant difference between the frequencies of nominalizations in PAK and ENS, these differences are not evenly distributed across the two corpora. In other words, nominalizations ending in suffix -tion were 'overused' in PAK whereas nominalizations ending in suffix -ment were 'overused' in ENS. This clearly shows some lexical preferences shown by Pakistani undergraduates and English native speakers. One of the significant contribution of this study is the starting point it provides for more diverse and in depth studies on nominalizations in the context of Pakistani Academic English. Further qualitative investigations focusing on syntactic patterns in which nominalizations occur in Pakistani Academic English may be a plausible extension of this study. Moreover, a comparative study of nominalizations in other similar Pakistani corpus maybe carried out.

An immediate implication of this study is for the teachers of English language, teaching Pakistani undergraduates argumentative essays. Teachers may use this research to inform their teaching practice to explicitly teach nominalizations and focus on developing the 'underused' or 'NOM-ment' nominalizations to Pakistani students. Similarly, 'NOM-tion' that are 'overused' by Pakistani students may also be taught carefully so that the students will be able to use them in a more effective manner. Additionally, material developers may use the findings to create more suitable teaching manuals and work books for teaching nominalization in argumentative essays.

Another benefit of this research maybe seen in the 'exclusion list' that was generated after applying the 'exclusion criteria' and this list may be used by corpus software program developers to develop a coding system or even a new software that automatically exploits the 'exclusion list' as a

stop list to tag nominalizations more accurately, especially for the researches who define nominalizations as Biber (1988) did.

References

[1] Abdulaziz,M. , Mahmdood,A.M. & Azhar,M.(2016) .Variation in learner's argumentative essays – a multi-dimensional comparative analysis. Science International 28 (4).

[2] Azher, M., Mehmood, M. A., & Shah, S. I. (2017). Linguistic Variation across Research Sections of Pakistan Academic Writing: A Multidimensional Analysis. International Journal of English Linguistics, 8(1), 30. doi:10.5539/ijel.v8n1p30

[3] Baratta ,A. M. (2010). Nominalization development across an undergraduate academic degree program. Journal of Pragmatics(42) 1017–1036

[4] Biber, D. (1988). Variation across speech and writing. Cambridge University Press.

[5] Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. Cambridge: Cambridge University Press.

[6] Biber, D., & Gray, B. (2013). Nominalizing the verb phrase in academic science writing. The verb phrase in English: Investigating recent language change with corpora, 99-132.

[7] Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman Grammar of Spoken and Written English. Essex: Pearson Education.

[8] Hinkel, E. (2002). Second language writers' text. Mahwah, NJ: Lawrence Erlbaum

[9] Ishikawa, S. (2018). The ICNALE edited essays; A dataset for analysis of L2 English learner essays based on a new integrative viewpoint. English Corpus Studies, 25, 117-130.

[10] Jalilifar, A., Alipour, M., & Parsa, S. (2014). Comparative study of nominalization in applied linguistics and biology books. Research in Applied Linguistics, 5(1), 24-43.

[11] Jalilifar, A., White, P., & Malekizadeh, N. (2017). Exploring nominalization in scientific textbooks: A cross-disciplinary study of hard and soft sciences. International Journal of English Studies, 17(2), 1-20.

[12] Kazemi, S. A. (2015). A Comparative Study of the Use of Nominalization in Articles Written by Native and Iranian Authors in English Teaching Discipline. The Iranian EFL Journal. Volume 11 Issue 2, 12(1), 149.

[13] Liardét, C. L., & Black, S. (2020). Trump vs. Trudeau: Exploring the power of grammatical metaphor for academic communication. Journal of English for Academic Purposes, 45, 100843. doi.org/10.1016/j.jeap.2020.100843

[14] Nausa, R. (2019). 'The Language of Oral Presentations Given by PhD Researchers in an EAP Class: Level of Performance and Disciplinary Differences'. Ph.D., University of Birmingham. http://etheses.bham.ac.uk/id/eprint/8883

[15] Sarani, A., & Talati, A. (2015). Nominalization in the Writing of Novice vs. Experienced Members of Academic Communities: A Comparative Discourse Analysis. Iranian Journal of Applied Language Studies, 7(2), 205-232. doi: 10.22111/IJALS.2015.2678

[16] Yoon, C. (2018). Nominalization in Korean EFL Learners' Argumentative Writing: A Comparative Study of Distribution and Use. The New Studies of English Language & Literature (69) 249-274.

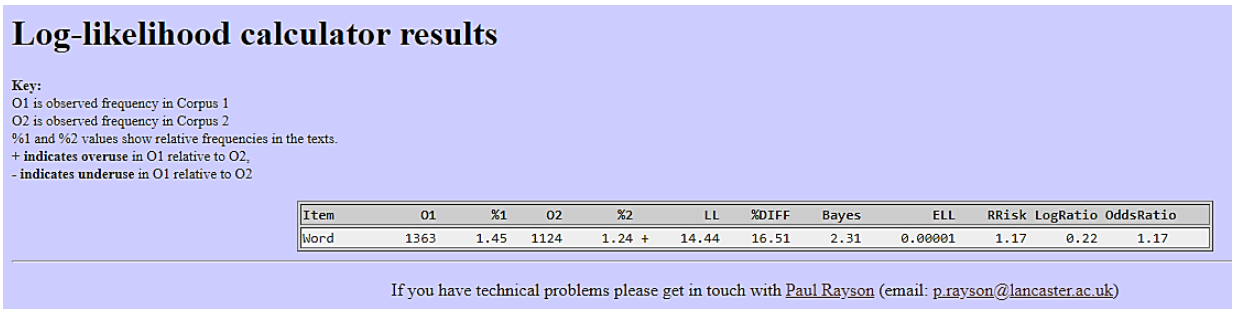
Appendix A

List of Words Excluded from the Analysis Based on the Exclusion Criteria(3.2.1)

Words not Derived from Adjectives (Adj) or Verbs(V)
NOM-tion; action, caution ,condition, function, institution, mention, nation, option, population, portion, position, question, recreation, station, transition, tuition **NOM-ment;** ailment, apartment, department, detriment, element, environment, implement, inclement, moment, supplement **NOM-ity;** capacity, charity, entity, facility, necessity, quality, university, utility ,vicinity **NOM-ness;** business
Concrete Nouns
partition, condition (air conditioner) , ornaments, instruments, compartment, harness

Appendix B

Raw Results of LL (and BIC) of Nominalizations in PAK with Respect to ENS



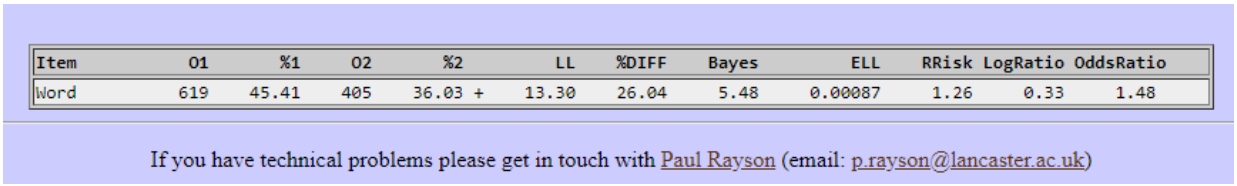
Log-likelihood calculator results

Key:
O1 is observed frequency in Corpus 1
O2 is observed frequency in Corpus 2
%1 and %2 values show relative frequencies in the texts.
+ **indicates overuse** in O1 relative to O2,
- **indicates underuse** in O1 relative to O2

Item	O1	%1	O2	%2	LL	%DIFF	Bayes	ELL	RRisk	LogRatio	OddsRatio
Word	1363	1.45	1124	1.24 +	14.44	16.51	2.31	0.00001	1.17	0.22	1.17

If you have technical problems please get in touch with Paul Rayson (email: p.rayson@lancaster.ac.uk)

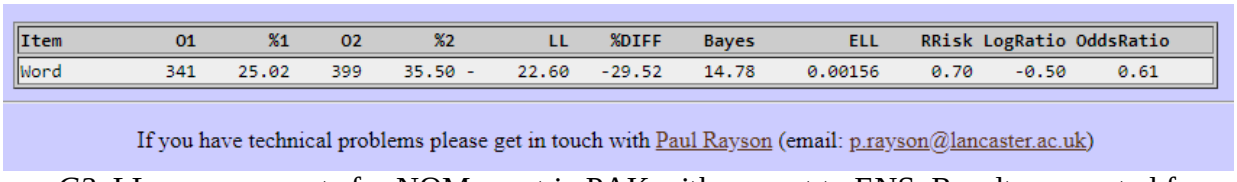
Figure G1. LL and BIC measurements on nominalizations in PAK with respect to ENS. Results generated from the online log likelihood calculator; http://ucrel.lancs.ac.uk/llwizard.html



Item	O1	%1	O2	%2	LL	%DIFF	Bayes	ELL	RRisk	LogRatio	OddsRatio
Word	619	45.41	405	36.03 +	13.30	26.04	5.48	0.00087	1.26	0.33	1.48

If you have technical problems please get in touch with Paul Rayson (email: p.rayson@lancaster.ac.uk)

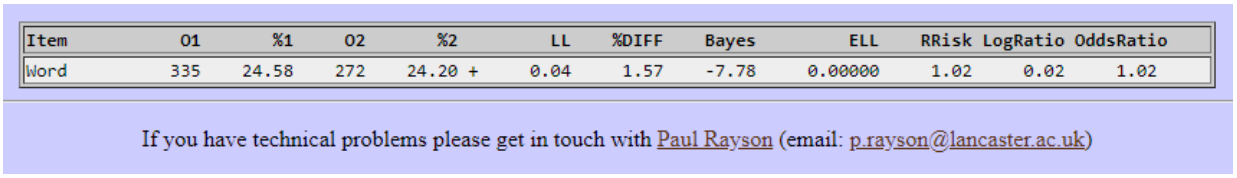
Figure G2. LL measurements for NOM-tion in PAK with respect to ENS. Results generated from the online log likelihood calculator; http://ucrel.lancs.ac.uk/llwizard.html



Item	O1	%1	O2	%2	LL	%DIFF	Bayes	ELL	RRisk	LogRatio	OddsRatio
Word	341	25.02	399	35.50 -	22.60	-29.52	14.78	0.00156	0.70	-0.50	0.61

If you have technical problems please get in touch with Paul Rayson (email: p.rayson@lancaster.ac.uk)

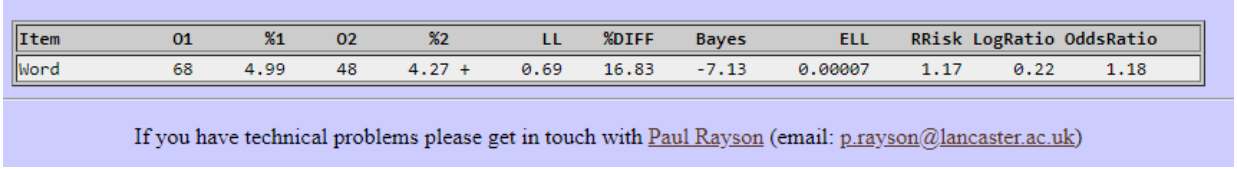
Figure G3. LL measurements for NOM-ment in PAK with respect to ENS. Results generated from the online log likelihood calculator , http://ucrel.lancs.ac.uk/llwizard.html



Item	O1	%1	O2	%2	LL	%DIFF	Bayes	ELL	RRisk	LogRatio	OddsRatio
Word	335	24.58	272	24.20 +	0.04	1.57	-7.78	0.00000	1.02	0.02	1.02

If you have technical problems please get in touch with Paul Rayson (email: p.rayson@lancaster.ac.uk)

Figure G4. LL measurements for NOM-ity in PAK with respect to ENS. Results generated from the online log likelihood calculator , http://ucrel.lancs.ac.uk/llwizard.html



Item	O1	%1	O2	%2	LL	%DIFF	Bayes	ELL	RRisk	LogRatio	OddsRatio
Word	68	4.99	48	4.27 +	0.69	16.83	-7.13	0.00007	1.17	0.22	1.18

If you have technical problems please get in touch with Paul Rayson (email: p.rayson@lancaster.ac.uk)

Figure G5. LL measurements for NOM-ness in PAK with respect to ENS. Results generated from the online log likelihood calculator , http://ucrel.lancs.ac.uk/llwizard.html