

Compiling Literary Academic Word List (LAWL): A Corpus Analysis of English Papers Published by the HEC Recognized Pakistani Journals

Rana Kashif Shakeel

Department of Humanities & Linguistics, University of Agriculture

Dr. Muhammad Kamal Khan

Department of English & Applied Linguistics, Allama Iqbal Open University

Abstract

The present study aims at compiling a Literary Academic Word List (LAWL) based on the HEC recognized Pakistani journals of English literature published from 2010 to 2019. A corpus consisting 215734 words with 16362 word forms was manually filtered to extract the most frequently used vocabulary items in 40 literature based research papers. The acquired wordlist (LAWL) of 766 words with a high frequency list (HFLAWL) consisting 20 words were chosen for critical analysis and discussion, considering the short and long contexts of the concordance lines and the external sources like key literary works. The results indicated that the selected timeframe was the period of frequent application of literary theories and critical approaches in Pakistani academic and research circles. The findings further stress the need for the compilation of subject specific word lists not only for literature but also for linguistics and other branches of knowledge. Finally, it is hoped that the availability of LAWL is going to be a significant contribution to the general process of learning English and to the specific process of exploring literary vocabulary in a robust manner.

Keywords: Context, HFLAWL, LAWL, lexical items, literary vocabulary, Pakistani journals

1. Introduction

The present research aims at compiling a *Literary Academic Word List (LAWL)* based on research articles published in HEC recognized Pakistani journals from 2010 to 2019. Research articles, write Khani and Tazik (2013), “[have become] a conspicuous tool for transmission of knowledge among scientists and researchers” (p. 209). Although research articles are comparatively short documents, their vocabulary is highly academic, genre oriented, technical and pithy. The genre of research articles, therefore, has been selected for the current study on literary vocabulary.

Reading and writing associated with vocabulary and word lists in the foreign language seem a baffling task for those who encounter with learning and teaching of a foreign or second language e.g., English in Pakistan. Chen and Ge (2007) endorse the preceding statement as, “[this] is also true for those who are learning a foreign language for academic purpose” (p. 503). The challenges related to reading and writing faced by the language users in their academic fields do not only relate with their subject/discipline specific knowledge but also with their language used for academic purposes (p. 503), and, specifically, for writing and reading about literary genres. As a matter of fact, the problem lies with the possession and usage of suitable vocabulary (Shaw, 1991). Chen and Ge (2007) acknowledge and explain Shaw’s point of view in

a more detailed manner by categorizing the problem of the unfamiliarity with the set of special word items (p. 503). These special words may either be literary vocabulary items or general academic vocabulary items. Coxhead (2000) explores the same issue by proclaiming that the users possess more vocabulary related to their subject than the general academic ones. She analyzes both of the categories by separating subject oriented vocabulary from the academic vocabulary. However, this area has not been well discussed by the critics so far. Nonetheless, more familiarity with the genre or subject related vocabulary and lesser expertise in academic vocabulary cause a great deal of difficulty for not only students but also for academicians. Yang (2015) gives the solution of the problem highlighted by Coxhead (2000) and proposes practicing the combination of general vocabulary with subject oriented vocabulary which might cover 85% of academic texts. The same idea proposed by Yang (2015) is being used in the present research by enlisting both the academic and literary vocabulary (subject specific vocabulary) in *LAWL* (See Appendix I).

Nation (2001) categorizes the academic words in terms of frequency of use by separating different categories such as *high frequency words*, *academic words*, *technical words*, and *low frequency words*. *High frequency words* are frequently used in colloquial conversation, but in all types of written scripts, they are the running words whereas *technical words* are linked with the specialized fields. Keeping in view the definition of *high frequency words*, we may observe that academic words ‘account for a relatively high proportion of running words in all academic texts’ (Chen & Ge, 2007, p. 503). In the same pattern, literary vocabulary belongs to *high frequency words category* as well as *technical words category*. The major purpose of the current study, therefore, is to establish a literary words list based on various literary genres, literary criticism and literary theories, as well as critical approaches being the ingredients of English research articles. The research, in the light of the definitions of *high frequency* and *technical words/subject oriented vocabulary* (literary words), describes which literary genres (novel, drama, poetry) and theories were more prevalent during these ten years (i.e., 2010-2019) in the HEC recognized Pakistani research journals of English (as a subject).

2. Literature Review

LAWL is a useful source for both English for Academic Purposes (EAP) instructors and learners, and corpus linguistic researchers. It was developed first by Averil Coxhead for her M.A. thesis. She provided the list of 570 headwords, which according to Maswana, Kanamaru and Tajino (2013), “[is] common to research papers across a broad range of disciplines” (p. 72). Kokkinakis, Skoldberg, Henriksen, Kinn and Johannessen (2012), and Vongpumivitch, Huang and Chang (2009) presented a more detailed study of Coxhead’s *AWL* (2000). It comprises about ‘400 texts from a range of academic articles, text-books and course books’ (Kokkinakis et al., 2012, p. 564). Furthermore, both the aforementioned sources explore that the size of corpus taken by Coxhead was 3.5 million tokens and the sub-corpora covered for the list were taken from the four areas—science, law, commerce and arts. According to Table-1 given by Kokkinakis et al. (2012), *AWL* covers 29 sub-categories, however, it lacks many important categories like literature, literary criticism, and literary theories. In the same research project, the idea of word family has been linked with the concept of *headwords*. Kokkinakis (2012) also points out the limitations of *AWL* and raises an objection on the word list developed by Coxhead (2000) by mentioning that she did not provide sufficient definitions or examples for further elaboration of her word list and ignored to see or describe the linkage between *AWL* and the

related disciplines. The lack of literary categories in *AWL* and less elaboration of word items (Kokkinakis et al., 2012) and enlisting only the headwords noticed by Maswana, Kanamaru and Tajino (2013) are the main research gaps to be filled in by this research at a micro level.

Despite all insufficiencies, *AWL* is still considered a touchstone design in the field of corpus linguistics. Other corpus linguistic researchers have developed their own word lists based on academic writings—research theses, articles, papers, and their abstracts by replicating, modifying and advancing the content and methods of *AWL*. Thongvitit and Thumawongsa (2017), and Farjami (2013) did research on the abstracts of research articles. The former selected the articles in the field of English as a Foreign language (EFL) written by Thai writers whereas the later did research on the abstracts of applied linguistics articles. Based on forty articles published from 2010 to 2015, the developed wordlists focused on the correct and incorrect use of lexical and grammatical collocations with the help of *AntConc*, *TagAnt*, and *The Oxford Collocations Dictionary for Students of English*, brought out in 2009 (Thongvitit & Thumawongsa, 2017, p. 752). On the other hand, Farjami (2013) explored the articles of applied linguistics published in different research journals from 2005 to 2011. The sources used by Farjami are *AWL*, *GSL*, *BNC*, and *Compleat Lexical Tutor*. The researcher delimited the study to the 100 most frequent words found in the abstracts of articles of applied linguistics. Further, it also endeavors to find out the share of 100 words in *AWL*, *GSL* etc. and highlights the insufficiency of the reference sources. The results show that prepositions stand at the top of the frequency, which is in line with the findings of Maswana, Kanamaru and Tajino (2013). However, it is also to be noted that most of the researches do not include prepositions in the word lists. In 2014, Jahangard et al. worked on a corpus of 400 research articles of *Hard Sciences* and analyzed the percentage of overlapping words between the source corpus and the reference sources (i.e. the word lists *AWL* and *GSL*). The researchers analyzed the data in two ways. First, the whole corpus was compared with the reference sources, and secondly, the sub-corpora (based on the articles of each category) were compared with the reference sources. Reference sources, here, have been referred to as *AWL* and *GSL*. After the discussion of the results, they pinpoint “[the] consistent need in *ESP* and *EAP* for specific corpora, to which the students can refer to check their language productions...” (p. 603). However, the students are very unlikely to refer to a corpus, because doing so requires very specific knowledge. On the other hand, Kokkinakis et al. (2012) gave a rather different trend to the process of vocabulary building. They, unlike the studies discussed above, proposed to utilize *AWL* for developing academic word lists for Swedish, Norwegian and Danish languages (p. 563).

Maswana, Kanamaru and Tajino (2013) applied the same method of Kokkinakis et al. (2012) on the corpus data ‘English expressions across 15 disciplines’. They developed a general corpus claiming to encompass all disciplines to generalize their results. The research aimed at the exploration of one-word and four-words expressions out of 20000 words (p. 71). Unlike, the results mentioned by Maswana, Kanamaru and Tajino (2013), the present research aims to provide a field specific world list. Moreover, Maswana, Kanamaru and Tajino (2013) also refer to the use of Coxhead’s *AWL*, and *GSL* as reference source by Michael West and the earlier researchers (p. 72-73). In their analysis, they used *The Kyoto University Academic Corpus* established in 2008 as a language source for teaching academic writing (p. 76). Mozaffari and Moini (2014) also did work on the corpus of 1.7 million words based on the research articles of *Education*. The research displays that 84115 words were matching with *AWL* word forms. Here, the frequencies ranged from the words – ‘Research’ (N=4767) to “Prior” (N=396).

Esfandiari and Moein (2015) worked on a project “A Corpus-driven Food Science and Technology Academic Word List”, in which they included “1421 research articles randomly selected from 38 journals across five sub-disciplines in *FST*” (p. 131). They used the criteria of frequency and range to develop *FSTAWL*. *EAP* and *AWL* have been cited as the reference sources. In this research, difficulty lies with the relationship between the *academic vocabulary* and *technical vocabulary*. After discussing the word lists—*high frequency vocabulary*, *low frequency vocabulary*, *technical vocabulary* and *academic vocabulary*, they developed a corpus of 465,244,4 words. Word families as found in the *AWL* were engaged to “identify frequency and specialized occurrences...” (p. 142). The findings of the research were delimited to 30 most frequent academic words in *FSTAWL*. The headwords from “Use” (N=27880) to “Food” (N=7227) consist 30 words with highest frequencies. By replicating the method used by Farjami (2013), Mozaffari and Moini (2014) and Esfandiari and Moein (2015), the acquired word list (See Appendix 1) may also be used as a reference source for literary studies.

Academic phraseology, frequency and distribution are the other issues concerning with the process of developing vocabulary lists and the role of *AWL*. Vongpumivitch, Huang and Chang (2009) presented a comparative study of the frequencies of *AWL* and *non-AWL* content words in the corpus of applied linguistics research papers. Unlike Farjami (2013) who used different reference corpora, they presented a proportion of words covered by *AWL* in *ALC* developed by them. The results show that *AWL* covers 11.17% of the words of *ALC* (Vongpumivitch, Huang & Chang, 2009, p. 36). In contrast with Farjami (2009), this research was delimited to the content words only. On the other hand, the results of *non-AWL* words showed that there were 128 *non-AWL* content word forms ‘that occur at least 50 times in *ACL*...collectively those 128 words occur 43001 times [accordingly for 2.8% of *ALC*]’ (p. 37).

The frequency of lexical items in *WPS* (*Whole Paper Corpus*) was analyzed by Jahangard et al. (2014) which explored the ratio of individual frequencies of individual words. Chen and Ge (2007) display the top 20 frequencies of individual sections to demonstrate the results. The same approach has been used in the current study (See Table 1). Wang, Liang and Ge (2008) established a medical academic word list based on 32 subject areas of medicine and dentistry (pp. 445-446). The results demonstrated through statistical tables the specialized occurrences, range and frequency of each lexical item selected for the analysis. The researchers also attached an appendix of a word list of 623 vocabulary items used in medical sciences. The list consisted of both academic words and subject oriented words as suggested by Yang (2015). On the other hand, Pathan et al. (2018) presented their research on the vocabulary used in doctoral theses of major scientific disciplines. Although *AWL* is not a corpus (but a word list), the researchers, like many others, used *AWL* as reference corpus and drew their own results that show 550-word families (which occur more than 10 times in the corpus) tallying 96.49% with *AWL* (p. 282).

Taking a step forward, Vincent (2013) investigated academic phraseology through contribution of very frequent words. The researcher identifies “the most common phrases in academic English” (p. 44). The results and discussions focus on the chains or lexical bundles and show them through various tables (pp. 48-53). The same pattern with somehow different subject matter i.e. *EAP* has also been adopted by Clouston (2013) and Aluthman (2017). They engaged word lists, as has also been described by Vincent (2013, p. 53) for the purpose of pedagogy of English. A very important aspect of Clouston’s (2013) research is the diachronic history and importance of word lists and their expanding scope with the passage of times. Keeping in view

the day to day increasing importance of word lists, he suggested the usage of two uncommon lists—*First 100 Spoken Collocations* by Shin and Nation (2008) and *Phrasal Expression Lists* by Martiner and Schmitt (2012) in addition to *GSL* by West (1953), *UWL* by Xue and Nation (1984) and *AWL* by Coxhead (2000). The first two lists have been introduced because of the insufficiency of the latter three lists. *First 100* and *PHRASE List* contain spoken collocations and discourse markers respectively. They seem to fill the gap left behind by West, Xue and Nation and Coxhead. He also cites eight other less known lists based on different genres—*BWL1* (Business Word List-1, 2007), *SWL* (Science Word List, 2007), *MAWL* (Medical Academic Word List, 2008), *Agrocorpus List* (2009), *BEL* (Basic Engineering Word List, 2009a), *NWL* (Newspapers Words List, 2009), *TWL* (Theological Word List, 2010), and *BWL2* (Business Word List-2, 2011a). Although the given list ignores the literary academic corpus, the paper demonstrates how these word lists may help us in the field of pedagogy of English. On the other hand, Aluthman (2017) compiled an *OPEC* word list, and a corpus-based lexical analysis was done.

The research works cited above identify the absence of a subject specific (*Literary Academic Word List*). Lessard-Clouston's (2013) presentation of subject specific corpora, Wang, Liang and Ge's (2008) work on the vocabulary of medical sciences, Esfandiari and Moein's (2015) compilation of *FSTAWL*, Jahangard et al.'s (2004) input to the lexical items of hard sciences, Nation's (2001) categorization of the academic words, and Yang's (2015) proposed list of vocabulary—all are the ready precedents of subject specific academic word lists. This review also indicates that no remarkable work is available on the literary academic vocabulary. So, the present research focuses on compiling a word list based on the literary research articles published in Pakistani research journals of English. It also aims at describing first 20 words (*HFLAWL*) to relate it with the current research trends in English literature in Pakistan. The students, teachers and literary critics may benefit from *LAWL* by incorporating the high frequency vocabulary in their literary writings. The research endeavors to answer the following questions:

1. What are the most frequent academic words in the *LAWL* of the HEC recognized Pakistani journals of English?
2. What do these *HFLAWL* (*high frequency literary and academic words*) suggest?

3. Research Methodology

The step by step procedures of the study are given in the following lines:

3.1. Data Collection and Sampling

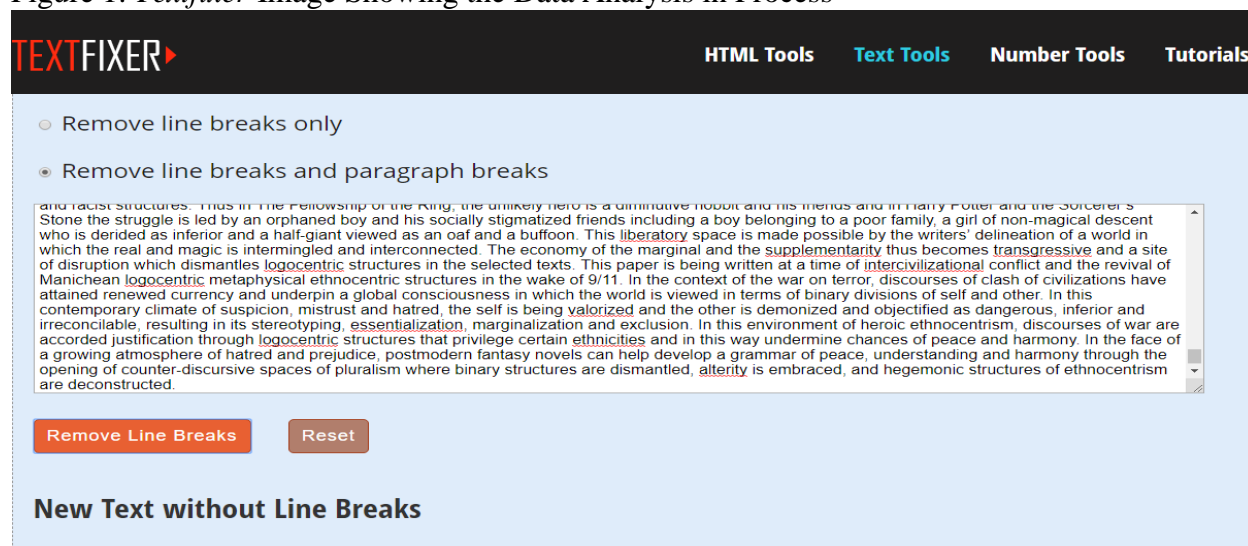
Data collected for this research comprises of the HEC recognized Pakistani research journals delimited to the category of *Arts and humanities*. This category has further many sub-categories, of which the journals relevant to English literature were selected for the study. The journals related to the subject of English literature are eight in number. The total accessible number of articles from these eight journals was 119 that were downloaded from the online archives of the public sector universities. Techniques of purposive and feasibility sampling were engaged to meet the needs of the research. Forty research articles containing a good number of words for feasible generalization of the results were selected to develop the corpus of 215734 words with

16362 word forms. All function and content words irrelevant to literary field were removed from 16362 word forms. After manual filtration, 766 words applicable to literary studies (already categorized) were found as literary vocabulary enlisted as *LAWL*. The list is appended with the paper as Appendix-I.

3.2. Corpus Establishment and Data Processing

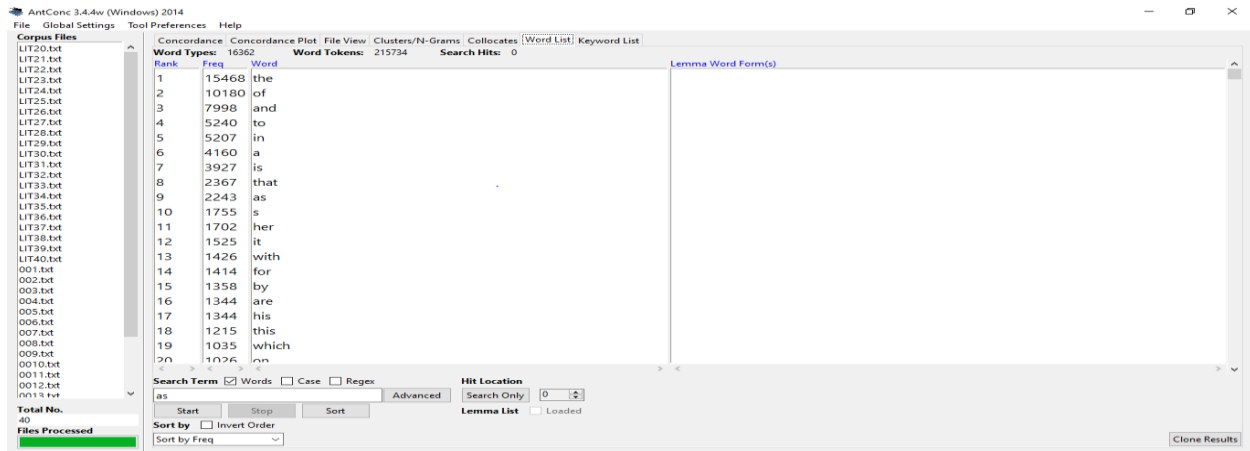
The acquired whole list named *LAWL* and *HFLAWL* (consisting of 20 highest frequency words) taken from 40 research articles written by Pakistani researchers were included in the study. For normalization and standardization, titles of the articles, bibliography, charts, visuals and diagrams were removed. After this, unnecessary spaces were also removed by online software *Textfixer* (shown in Figure 1).

Figure 1: *Textfixer* Image Showing the Data Analysis in Process



In the next step, the cleaned files were converted into *TXT* format. Finally, all the files were processed by the software *AntConc*. When all files were uploaded, the software was commanded to make a word list (shown in Figure 2).

Figure 2: *AntConc* Image Showing the Word List as Processed



After processing the text data through *Textfixer* and *AntConc*, the word lists were analyzed, and the respective frequencies were taken for the selected literary terms and word lists. On the basis of statistics taken from this step, the high frequency academic word lists were counted. The list of first twenty most frequently words is given in the following table:

Table 1: High Frequency Academic Word List

Ranking	Frequency	Word	Ranking	Frequency	Word
1	462	Other	11	252	Identity
2	458	Power	12	252	Man
3	399	Social	13	249	Time
4	372	Language	14	236	Characters
5	359	Discourse	15	226	Culture
6	354	Cultural	16	223	Self
7	334	Women	17	212	Reality
8	302	Story	18	208	Novel
9	289	Society	19	207	Literature
10	278	Text	20	193	Literary

No reference corpus has been referred but a few keywords on literary studies, short and long contexts of the nodes, and exemplary word clusters have been cited to strengthen the arguments while discussing and analyzing *HFLAWL*. The findings and discussion are given in the next section.

4. Findings and Discussion

Table-1 shows twenty words with the highest frequencies and most frequently used word being *other* with N=462. The concordance of the word with the short and long contexts highlight that the word has been used as noun, pronoun, verb, adverb and adjective in the corpus. But in terms of literary terminology, it signals the presence of binaries between the “self” and “other”, a key content of literary texts has also been accounted in various books of theory, especially the texts on postcolonial studies like *Orientalism* (2003) by Edward Said, *Postcolonial Studies: The Key Concepts* (2007) and *The Post-Colonial Studies Reader* (2003) by Bill Ashcroft, Gareth Griffiths and Helen Tiffin. The same rhetoric can be seen in one of the concordance lines as: “...subject as primitive, uncivilized, exotic, *other* binaries of East/West, Orient and Occident...”. In the same way, left and right sides of concordance lines making clusters “*other* colonies of Whites”, “the *other* become vengeful”, and “indispensable *other* suppressed” also provide the same themes.

The second ranked word in the *HFLAWL* is *power* with 458 hits. It has been used as noun in the whole corpus. The word engaged in the context of power-relation theories has mostly been associated with postcolonial studies and its applications on literature. Left and right words combined with the word *power* as “occupying”, “ruling and subjugation”, and “white” were the major instances taken from concordance lines of the corpus. Another exemplary concordance line shows the word *power* in relation with the word *other* as: “... its faith in a *power*[ful] play of eliminating the *other* as contender...”, which is debated upon by Ashcroft, Griffiths and Tiffin (2003) under the key terms “imperial *power*”, “knowledge and *power*” etc. and generates a relationship between *power* and *post-colonialism* as:

Such a situation simply reproduces the inequalities of imperial power relations. Post-colonial ‘theory’ has been produced in all societies into which the imperial force of Euro has intruded, though not always in the formal guise of theoretical texts (p. 2).

The word *social* with 399 hits having associations with other branches of knowledge and in particular with cultural, anthropological, and political studies has acquired the rank of the third highest frequency literary word in *HFLAWL*. Apart from its being a common noun category, it has become a highly literary academic word. Selden, Widdowson and Brooker (2005) endorse it as:

Most critics assumed, like Dr Johnson, that great literature was universal and expressed general truths about human life . . . [and] talked comfortable good sense about the writer’s personal experience, the social and historical background of the work, the human interest, imaginative ‘genius’ and poetic beauty of great literature... they must confront the problematical issues raised about ‘Literature’ and its social relations by major theorists in recent years. (pp. 1-4).

Left and right words engaged to make clusters with the lexical item *social* can be seen as: “... culture, *social*, regional, and historical generalizations...”, “*social* forces”, “physics and *social* forces”, “*social* space for women”, and “*social* injustice” represent literary theories as well as literary texts. Whereas the word *social* has the frequency of 399, the word *society* is with

the frequency of 354, hence making it 9th lexical item in the list. This lexical item has combinations with multiple fields of scholarship, cultural studies and literature in particular. Some of the examples in association with the word *society* are “sufi and cultural dimensions”, “local cultural aspects”, and “deeper cultural connotations”. From these examples, it looks like a natural adjustment in the literary articles of English literature. The concordance centered by the lexical item *language* comprises of 372 concordance lines, and combined with certain specific determiners reflect the influence of literary theories like postcolonialism, feminism, deconstruction, and postmodernism (Selden, Widdowson & Brooker, 2005, p. 18; Ashcroft, Griffiths & Tiffin, 2003, p. 7-55; Said, 2003, p. 22-25; Nayar, 2009, p. 7-30), and literary terminology like ‘symbolical’, ‘simple’ and ‘context’. Another word closer to the word “language” is “discourse”. It is at the 5th position in the ranking on the basis of its frequency. It is primarily associated with the postcolonial studies but it is also associated with other theories. The word clusters consisting of ‘written *discourse*’, ‘narrative *discourse*’, and ‘literary *discourse*’ refer to the field of literature.

The lexical item *women* with its headword *woman* has been employed by the critics in their research articles in abundance. In this analysis, the word *women* practiced more than its singular form belongs, in particular, to gender studies, feminism and cultural studies. The word with 334 hits occurs at 7th position in the *HFLAWL*. There may be several reasons behind this position, but the most significant reason is the production of feminist literature in great bulk and interest in feminist theory in the recent times. First, a few concordance lines show women as domestic, poor, passive, struggling and subjugated figure. In contrast, at certain points the clusters show them as ‘center to the plot’, ‘participation in public life’, and ‘liberal’. But the proportion between the two categories is uneven and is in the favor of the first.

The usage of the next word *story* as a noun with frequency (N=302) reflects its literal meanings i.e., narrative, account, plot etc. Clusters from the concordance prop up the very argument by providing the clusters like ‘elements in the *story*’, ‘case of *story* writing’, and ‘the Indian *story*’. So is the case with the word *text* with 278 hits which signifies to the literary studies, criticism and theory in the research articles. A *text* does mean wording of anything written or printed (*The Reader’s Digest Great Encyclopaedic Dictionary*, 1964, p. 914). In the light of the definition, almost all the nodes in the concordance refer towards the same commodity, i.e., literary texts, in the corpus. A concordance line itself endorses this point of view as: ‘a *text* is a composite picture of quotations’. Left and right collaborating words also support it as well as relate the word *text* with certain literary theories, especially formalism, structuralism, and Post-structuralism. But most of all, *text*, here means a literary text like novel, drama and poetry etc.

The word *identity* has a very wide scope in the field of literary studies. The total hits (N=252) of the very word inform us how the word *identity* has captured the attention of the authors and the critics of the literary work in the recent times. Moreover, it also reflects that modern literature picturizes the identity crises, or struggle for identity. So, it involves and represents one of the reoccurring themes of the modern literary studies. ‘European *identity*’, ‘Indian *identity*’, or ‘*identity* of Native America’ describe that Pakistani literary critics are doing work on the issue of identity frequently.

The word *man* with 252 hits has different connotations in different contexts. A few hits, perhaps, based on original texts unveil the metaphorical and symbolic layers of the word, but most of the concordance lines present *man* as a male human being. Clusters “a *man* is appreciated”, “a *man* may own a woman”, and “white *man*” are gender specific examples. However, metaphorical implications of the word can be observed as ‘a brave person, or a love’ in the clusters of the concordance. Even, the word in literary studies has been engaged as a common term for all humanity regardless of gender e.g. the symbol of “Gidh is *man*’s ethical collapse”

Three words *time* (N=249), *reality* (N=212), and *literature* (N=207) have been practiced in their pure literal sense. The lexical items *time*, *literature* and *reality* seen through their context have lesser chances of being used as metaphor. But in the case of the second word we can see that it is genre specific lexical item. The word *literature* is attached with “Pakistani”, “Pakistan” and “English” to make clusters. It also shows what type of literary works were chosen by the Pakistani literary critics during 2010-2019. Same treatment has been done with the literary genre *novel* (N=208), literary term *character* (N=236), and theory *culture* (N=226) by the critics. They have also been used almost adopted and adapted very like the word *literature*. As a matter of fact, the results of these four words usage remain probably identical in the corpus. But it is also to be noted that like the words *society* and *social*, the word *culture* has been employed by the critics in terms of post-colonial studies, feminism, and Marxism.

The word *self*, in fact a very important theoretical term appears in many different contexts. Used as noun and adjective, its main contribution refers to as a binary opposition of the term *others* that is the 1st ranked word in our HFALWL.

5. Conclusion

Focusing on our first research question (RQ-1), we compiled the *LAWL* (see Appendix-1) and the *HFLAWL* (see Table-1) from a corpus collected from 40 research papers published by leading Pakistani universities during 2010-19. The *LAWL* showed 766 words of different frequencies and the *HFLAWL* showed; *Other, Power, Social, Language, Discourse, Cultural, Women, Story, Society, Text, Identity, Man, Time, Characters, Culture, Self, Reality, Novel, Literature* and *Literary* with their respective ranking.

Subsequently, we used the data from the *LAWL* and *the HFLAWL* to address our RQ-2. Our findings from these lists and the subsequent discussion demonstrated that the stylistic choices, and theoretical and thematic trends prevailing in the contemporary writings of Pakistani literary critics were mainly at the backdrop of these words. Similarly, a significant part of the intellectual production during 2010-2019 was influenced by literary theories and criticism. Moreover, literary theories found at the first few positions were post-colonialism, cultural studies, feminism, language and structure-oriented theories and postmodernism. However, this variety of theories with their specific names did not occur in the first 50 frequencies, rather the key concepts had a good currency, recency and relevance. Another discouraging fact for the postcolonial critics was that the terms debated more in the classrooms and conferences were lesser used in the academic literary writings. For example; ‘hybridization’, ‘narratology’, ‘orientalism’ occurred 11 times in the list, whereas the keywords ‘ambivalence’, ‘feminist’, ‘subjugated’ did occur for 10 times. Many other well discussed terms like ‘Marxism’,

'abrogation', 'exoticism', 'neocolonial', 'existential' etc. occurred less than three times in the list. Overall, the wordlists *LAWL* and *HFLAWL* showed the current literary and academic trends within Pakistani research circles and the scholars working in the field of literary studies might take benefits from these lists in their academic writings.

6. Recommendations

In the light of this research, we recommend that:

- i. The design of research adopted in this research may be replicated and utilized by other researchers.
- ii. The research explains only 20 top frequency words out of 766. The other words may be taken by researchers for further exploration.
- iii. Further argumentative studies in favour or in rebuttal may also improve this area of study.

References

- Aluthman, E. S. (2017). Compiling an OPEC word list: A corpus-informed lexical analysis. *International Journal of Applied Linguistics and English Literature*, 6(2), 78.
- Ashcroft, B., Griffiths, G., & Tiffin, H. (2003). *Post-colonial studies: The key concepts*. Routledge.
- Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, 26(4), 502-514.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Esfandiari, R., & Moein, G. (2015). A corpus-driven food science and technology academic word list. *Issues in Language Teaching*, 4(2), 131-157.
- Farjami, H. (2013). A corpus-based study of lexical make-up of applied linguistics articles abstracts. *The Journal of Teaching Language Skills*, 5(2), 27-50.
- Jahangard, A. (2014). An investigation of academic word list (AWL) and general word list (GWL) in hard sciences' research articles: Meeting the rising demands in ESP. *The Journal of Teaching English for Specific and Academic Purposes*, 2(4), 603-614.
- Khani, R., & Tazik, K. (2013). Towards the development of an academic word list for applied linguistics research articles. *RELC Journal*, 44(2), 209-232.
- Kokkinakis, S., Skoldberg, E., Henriksen, B., Kinn, K., & Johannessen, J. B. (2012). *Developing academic word list for Swedish, Norwegian and Danish-A joint research project*. In *Euralex*.
- Lessard-Clouston, M. (2013). Word lists for vocabulary learning and teaching. *The CATESOL Journal*, 24(1), 287-304.
- Maswana, S., Kanamaru, T., & Tajino, A. (2013). Analyzing the journal corpus data on English expressions across disciplines. *The Journal of Asia TEFL*, 10(4), 71-96.
- Mozaffari, A., & Moini, R. (2014). Academic words in education research articles: A corpus study. *Procedia-Social and Behavioral Sciences*, 98, 1290-1296.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nayar. (2010). *Contemporary literary and cultural theory: From structuralism to Ecocriticism*. Pearson Education India.
- Pathan, H., Memon, R. A., Memon, S., Shah, S. W., & Magsi, A. (2018). Academic vocabulary use in doctoral theses: A corpus-based lexical analysis of academic word list (AWL) in major scientific disciplinary groups. *International Journal of English Linguistics*, 8(4), 282.
- The reader's digest great encyclopedic dictionary. (1964). The Reader's Digest Association. London.
- Said, E. W. (2003). *Orientalism*. Vintage.
- Selden, R., Widdowson, P., & Brooker, P. (2005). *A reader's guide to contemporary literary theory*. Pearson Education India.
- Shaw, P. (1991). Science research students' composing processes. *English for Specific Purposes*, 10(3), 189-206.
- Thongvitit, S., & Thumawongsa, N. (2017). A corpus-based study of English collocations found in the abstracts of research articles written by Thai EFL writers. *International Journal of Social Sciences and Humanity*, 7(12), 751-755.

- Vincent, B. (2013). Investigating academic phraseology through combinations of very frequent words: A methodological exploration. *Journal of English for Academic Purposes*, 12(1), 44-56.
- Vongpumivitch, V., Huang, J., & Chang, Y. (2009). Frequency analysis of the words in the academic word list (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33-41.
- Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*. 27(4), 442-458.
- Yang, M. (2015). A nursing academic word list. *English for Specific Purposes*. 37(1), 27-38.

Appendix-I
Academic Word List

Ranking	Frequency	Word
1	462	Other
2	458	Power
3	399	Social
4	372	Language
5	359	Discourse
6	354	Cultural
7	334	Women
8	302	Story
9	289	Society
10	278	Text
11	252	Identity
12	252	Man
13	249	Time
14	236	Characters
15	226	Culture
16	223	Self
17	212	Reality
18	208	Novel
19	207	Literature
20	193	Literary
21	182	Native
22	178	Words
23	177	Analysis
24	162	Dream
25	145	Subject
26	143	Writers
27	142	Colonial
28	125	Narrative
29	122	Work
30	121	History
31	120	Linguistic
32	115	Place

33	114	White
34	113	Meaning
35	112	Space
36	103	Discursive
37	103	Form
38	103	Reader
39	102	Concept
40	102	Western
41	100	Act
42	100	Individual
43	95	Political
44	94	International
45	94	Writing
46	91	Poetry
47	91	Role
48	90	Context
49	89	Modern
50	88	Representation
51	87	Experience
52	87	Patriarchal
53	87	Poet
54	87	Sentence
55	86	Fiction
56	86	Voice
57	85	Conscious
58	85	Structures
59	84	Marriage
60	83	Powerful
61	83	Theory
62	81	Muslim
63	80	Male
64	80	Unconscious
65	79	Critical
66	79	System
67	79	Understanding
68	78	Existence

69	76	Patriarchy	105	59	Values
70	76	Tribal	106	58	Book
71	75	Consciousness	107	58	Branding
72	74	Control	108	58	Superhero
73	71	Author	109	58	Truth
74	71	Magical	110	57	Poem
75	69	Sufi	111	56	Desire
76	68	Physical	112	56	Jungle
77	68	Readers	113	56	War
78	67	Binary	114	55	Difference
79	67	Perspective	115	55	Knowledge
80	67	Religious	116	54	Aspects
81	65	Historical	117	53	Interpretation
82	65	Image	118	53	Level
83	65	Postcolonial	119	53	Plays
84	65	Presence	120	53	Struggle
85	65	Read	121	53	Studies
86	65	Sign	122	53	Violence
87	65	Word	123	52	Parody
88	64	Community	124	52	Rich
89	64	Folk	125	52	Traditions
90	64	Question	126	51	Black
91	64	Speech	127	51	Colonized
92	62	Class	128	51	Moral
93	62	Concepts	129	50	Deconstruction
94	61	Approach	130	50	Nation
95	61	Gender	131	50	Symbols
96	61	Symbolic	132	49	Absence
97	60	Ego	133	49	Economic
98	60	Experiences	134	49	Imperial
99	60	Expressions	135	49	Inner
100	60	Reading	136	49	Postmodern
101	59	Psychological	137	49	Reference
102	59	Resistance	138	49	Sexual
103	59	Socio	139	48	Local
104	59	Structure	140	48	Practice

141	47	Communication	177	40	Pre
142	47	Contemporary	178	40	Traditional
143	47	Features	179	39	Interaction
144	46	Central	180	39	Politics
145	46	Earth	181	39	Psychic
146	46	Foe	182	39	Rights
147	45	Post	183	39	Style
148	45	Race	184	38	Comparative
149	45	Simulacra	185	38	Elements
150	44	Authority	186	38	Genre
151	44	Euro	187	38	Id
152	44	Patterns	188	38	Indigenous
153	43	Actions	189	38	Logic
154	43	Exercise	190	38	Model
155	43	Feminist	191	38	Perception
156	43	Movement	192	37	Contrast
157	43	Norms	193	37	Global
158	43	Oppression	194	37	Magic
159	43	Points	195	37	Religion
160	43	Professor	196	37	Resources
161	43	Psyche	197	37	Rewriting
162	43	Vision	198	37	Section
163	42	Anti	199	37	Series
164	42	British	200	37	Spiritual
165	42	Evil	201	37	Strategy
166	42	Forces	202	37	Textual
167	42	Marginalized	203	37	Tribes
168	42	Represent	204	37	Warrior
169	41	Challenge	205	36	Agency
170	41	Colonizers	206	36	Atmosphere
171	41	Issues	207	36	Comparison
172	41	Themes	208	36	Constructed
173	41	Version	209	36	Domestic
174	41	Written	210	36	Dominant
175	40	Brands	211	36	Structural
176	40	Intellectual	212	36	West

213	36	Write	249	30	Feelings
214	35	Framework	250	30	Humans
215	34	Complex	251	30	Hybrid
216	34	Construction	252	30	Origin
217	34	Country	253	30	Search
218	34	National	254	30	Single
219	34	Objects	255	30	Socially
220	34	Participants	256	30	Subjects
221	34	Protagonist	257	30	Systems
222	34	Researchers	258	29	Background
223	34	Wealth	259	29	Categories
224	34	Writings	260	29	Discipline
225	33	Civilization	261	29	Essential
226	33	Exist	262	29	Exploitation
227	33	Freedom	263	29	Fall
228	33	Ideology	264	29	Fight
229	33	Imperialism	265	29	Figures
230	33	Institutional	266	29	Hyper
231	33	Journey	267	29	Imagination
232	33	Narration	268	29	Logocentric
233	33	Standards	269	29	Master
234	33	Theoretical	270	29	Peace
235	33	Verbal	271	29	Popular
236	32	Analyzed	272	28	Circumstances
237	32	Portrayed	273	28	Code
238	32	Racial	274	28	Collective
239	31	Conflict	275	28	Globalization
240	31	Images	276	28	Intertextual
241	31	Justice	277	28	Market
242	31	Multiple	278	28	Poetic
243	31	Roots	279	28	Poor
244	30	Alterity	280	28	Setting
245	30	Art	281	28	Structuralism
246	30	Center	282	28	Superiority
247	30	Education	283	27	Acts
248	30	Ethnic	284	27	Basis

285	27	Creative	321	24	Universal
286	27	Dark	322	23	Applied
287	27	Dialogue	323	23	Article
288	27	Establish	324	23	Colonialism
289	27	Harmony	325	23	Counter
290	27	Ideological	326	23	Critics
291	27	Persona	327	23	Dominance
292	27	Phenomena	328	23	Fictional
293	27	Superego	329	23	Humor
294	26	Ancient	330	23	Imaginative
295	26	Course	331	23	Passive
296	26	Domination	332	23	Production
297	26	Facts	333	23	Rural
298	26	Laws	334	22	Challenging
299	26	Liberal	335	22	Contents
300	26	Narrator	336	22	Conversation
301	26	Positive	337	22	Empire
302	26	Scene	338	22	Ethnocentric
303	26	Worlds	339	22	Heritage
304	25	Development	340	22	Metaphysical
305	25	Exercised	341	22	Narratives
306	25	Lexical	342	22	Parts
307	25	Media	343	22	Philosophy
308	25	Stereotypes	344	22	Picture
309	25	Supplement	345	22	Psychology
310	25	Types	346	22	Rooted
311	24	Academic	347	22	Sex
312	24	Cognitive	348	22	Slave
313	24	Connection	349	22	Speaker
314	24	Criticism	350	22	Technological
315	24	Formation	351	22	Universe
316	24	Masculine	352	21	Accepted
317	24	Medium	353	21	Audience
318	24	Rule	354	21	Canonical
319	24	Sexuality	355	21	Corporate
320	24	Standard	356	21	Corporations

357	21	Definition	393	19	Situations
358	21	Depicts	394	19	Sources
359	21	Dimensional	395	19	Traits
360	21	Multi	396	19	Transformation
361	21	Notions	397	18	Argument
362	21	Oppressive	398	18	Assumptions
363	21	Presentation	399	18	Conventions
364	21	Primitive	400	18	Critique
365	21	Privileged	401	18	Feudal
366	21	Satire	402	18	Frequency
367	21	Speak	403	18	Freudian
368	21	Stylistic	404	18	Function
369	21	Utterance	405	18	Metropolitan
370	20	compared	406	18	Movies
371	20	Current	407	18	Mystical
372	20	Elite	408	18	Nationalism
373	20	Feminine	409	18	Subaltern
374	20	Hegemony	410	18	Suffering
375	20	Identify	411	18	Superior
376	20	Irony	412	18	Symbolically
377	20	Metaphor	413	18	Technology
378	20	Plot	414	18	Title
379	20	Rational	415	18	Tools
380	20	Referring	416	18	Web
381	20	Signified	417	17	Canto
382	20	Statement	418	17	Comic
383	20	Suggests	419	17	Customs
384	20	Translation	420	17	Deconstructive
385	19	Conscience	421	17	Essay
386	19	Culturally	422	17	Fantasy
387	19	Define	423	17	Field
388	19	Oppositions	424	17	Hegemonic
389	19	Partition	425	17	Implications
390	19	Principle	426	17	Influences
391	19	Realism	427	17	Inherent
392	19	Signifier	428	17	Masses

429	17	Method	465	15	Finding
430	17	Novelist	466	15	Government
431	17	Oppressed	467	15	Humanity
432	17	Paragraph	468	15	Intended
433	17	Parodical	469	15	Interpreted
434	17	Published	470	15	Liberation
435	17	Signifies	471	15	Objective
436	17	Societal	472	15	Progressive
437	16	Classic	473	15	Qualitative
438	16	Communicative	474	15	Racist
439	16	Comprehension	475	15	Syntactic
440	16	Construct	476	15	Topic
441	16	Contextual	477	15	Violent
442	16	Cosmopolitan	478	14	Agenda
443	16	Domain	479	14	Analytical
444	16	Dreamer	480	14	Dominated
445	16	Duration	481	14	Dynamics
446	16	Encounter	482	14	Epic
447	16	Equality	483	14	Ethical
448	16	Exotic	484	14	Independent
449	16	Gendered	485	14	Institutions
450	16	Memory	486	14	Intelligence
451	16	Myth	487	14	Interpretive
452	16	Philosopher	488	14	Key
453	16	Phrase	489	14	Literally
454	16	Portrayal	490	14	Marginal
455	16	Subjectivity	491	14	Opinion
456	16	Summary	492	14	Orient
457	16	Tale	493	14	Otherness
458	16	Uncivilized	494	14	Prose
459	16	Unconsciously	495	14	Scenario
460	16	Variety	496	14	Stereotypical
461	16	Vocabulary	497	14	Superficial
462	16	Vulnerable	498	14	Suppressed
463	15	Contrapuntal	499	14	Supremacy
464	15	Devices	500	14	Techniques

501	14	Tragic	537	12	Agents
502	14	Victims	538	12	Classical
503	14	Weapons	539	12	Constituted
504	13	Analogies	540	12	Economy
505	13	Capital	541	12	Heroic
506	13	Characteristics	542	12	Identified
507	13	Colonization	543	12	Intuition
508	13	Eurocentric	544	12	Loneliness
509	13	Explicit	545	12	Loses
510	13	Grammar	546	12	Manifest
511	13	Grammatical	547	12	Manners
512	13	Hero	548	12	Meaningless
513	13	Hierarchy	549	12	Normative
514	13	Ironically	550	12	Para
515	13	Manipulation	551	12	Passage
516	13	Military	552	12	Revolution
517	13	Mood	553	12	Satirical
518	13	Mysterious	554	12	Suppression
519	13	Narrates	555	12	Symbolism
520	13	Neo	556	12	Temporal
521	13	Pages	557	12	Translated
522	13	Parallel	558	12	Trend
523	13	Periphery	559	12	Unity
524	13	Philosophical	560	11	Artificial
525	13	Pragmatic	561	11	Barbaric
526	13	Principles	562	11	Bourgeois
527	13	Projected	563	11	Civilized
528	13	Reaction	564	11	Depiction
529	13	Regime	565	11	Drama
530	13	Regional	566	11	Educated
531	13	Regions	567	11	Exploiters
532	13	Resisted	568	11	Hybridization
533	13	Skin	569	11	Narratology
534	13	Spoken	570	11	Orientalism
535	13	Super	571	11	Paradigm
536	12	Abstract	572	11	Subjective

573	11	Symbolizes	609	9	Structured
574	11	Trilogy	610	9	Territory
575	10	Ambivalence	611	9	Terror
576	10	Colonialist	612	9	Thematic
577	10	Colonies	613	9	Theorists
578	10	Connotations	614	8	Capitalistic
579	10	Cosmopolitanism	615	8	Diasporic
580	10	Feminism	616	8	Dismantle
581	10	Idiom	617	8	Euroamerican
582	10	Imperialist	618	8	Globe
583	10	Injustice	619	8	Glocalization
584	10	Marginalized	620	8	Parodic
585	10	Masculinity	621	8	Signification
586	10	Meaninglessness	622	8	Signify
587	10	Multiplicity	623	8	Subjected
588	10	Mysteries	624	8	Subordinate
589	10	Narrativity	625	7	Allegory
590	10	Narratological	626	7	Ballad
591	10	Savage	627	7	Base
592	10	Subjugated	628	7	Corpus
593	10	Sufism	629	7	Dictators
594	9	Anglo	630	7	Epistemology
595	9	Critically	631	7	Imaginary
596	9	Dialectics	632	7	Inhuman
597	9	Dialogic	633	7	Postmodernism
598	9	Discussions	634	7	Pseudo
599	9	Dismantling	635	7	Psychoanalysis
600	9	Geographical	636	6	Alienation
601	9	Humour	637	6	Allegorical
602	9	Hyperreal	638	6	Ambivalent
603	9	Literal	639	6	Arbitrary
604	9	Metaphysics	640	6	Archetypal
605	9	Migrants	641	6	Decolonization
606	9	Racism	642	6	Disorder
607	9	Spatial	643	6	Displacement
608	9	Structuralist	644	6	Episode

645	6	Ethnicity	681	4	Empirical
646	6	Figurative	682	4	Idealism
647	6	Mimicry	683	4	Intertextuality
648	6	Modernity	684	4	Metamorphosis
649	6	Mythology	685	4	Metarepresentation
650	6	Nationality	686	4	Metonymy
651	6	Tragedy	687	4	Multiculturalism
652	5	Anthropology	688	4	Pluralistic
653	6	Heteroglossia	689	4	Readership
654	5	Homeland	690	4	Readings
655	5	Homosexual	691	4	Regressive
656	5	Humanities	692	4	Ridiculous
657	5	Hybridity	693	4	Romanticized
658	5	Margin	694	4	Semantic
659	5	Marginalization	695	4	Semiology
660	5	Meta	696	4	Semiotic
661	5	Migration	697	4	Syntax
662	5	Negotiate	698	4	Systematic
663	5	Negotiation	699	4	Systematically
664	5	Ontology	700	4	Systemic
665	5	Orientalists	701	4	Terminology
666	5	Pluralism	702	4	Theological
667	5	Pragmatics	703	4	Theoretically
668	5	Radical	704	4	Theorized
669	5	Reclamation	705	4	Therapeutic
670	5	Repressive	706	3	Activists
671	5	Romantics	707	3	Actor
672	5	Simile	708	3	Afrocentric
673	5	Subtexts	709	3	Americana
674	4	Analogy	710	3	Americanism
675	4	Anthropological	711	3	Analyst
676	4	Chronology	712	3	Analytic
677	4	Counterproductive	713	3	Anglian
678	4	Deceit	714	3	Anglophone
679	4	Democratic	715	3	Antagonistic
680	4	Demographic	716	3	Dialect

717	3	Epitome	753	2	Interpretative
718	3	Hermeneutics	754	2	Interpreter
719	3	Monopoly	755	2	Marxism
720	3	Motif	756	2	Marxist
721	3	Multilingual	757	2	Masterpiece
722	3	Narratologists	758	2	Neocolonial
723	3	Negro	759	2	Prototype
724	3	Ritual	760	2	Rationalist
725	3	Sociological	761	2	Refugees
726	3	Subordination	762	2	Storytellers
727	3	Textualization	763	2	Surrealist
728	3	Textulaized	764	1	Abrogation
729	3	Theatre	765	1	Ahistorical
730	3	Thesis	766	1	Ahistoricity
731	2	Aboriginal			
732	2	Accents			
733	2	Allegoric			
734	2	Allegorically			
735	2	Allusion			
736	2	Amorphous			
737	2	Analysts			
738	2	Analytically			
739	2	Contextualize			
740	2	Corrective			
741	2	Disillusionment			
742	2	Etymologically			
743	2	Existential			
744	2	Exoticism			
745	2	Folklores			
746	2	Homogenization			
747	2	Homogenous			
748	2	Hyperbole			
749	2	Hypnagogic			
750	2	Hypnopompic			
751	2	Hypocritical			
752	2	Interior			