A Corpus-Based Analysis of the Lexical and Grammatical Properties of Arabic Abstracts in Social Sciences Articles

Alaa’ Mohammad Smadi
Basic Sciences Department, Balqa’ Applied University, Salt, Jordan
E-mail: alsmadi_alaa88@yahoo.com

Received: October 9, 2020  Accepted: November 17, 2020  Published: November 24, 2020
doi:10.5296/ijl.v12i6.17849  URL: https://doi.org/10.5296/ijl.v12i6.17849

Abstract
This study aims to analyze quantitatively and qualitatively Arabic journal articles’ abstracts written within the field of social sciences. It mainly aims to analyze the lexical and grammatical qualities of the abstracts in the five academic disciplines; Economy, Geography, Psychology, Sociology, and Law. To achieve the goal of the study, a corpus consisting of 500,000 words was collected from various well-known Arabic journals, and then it was divided into five sub-corpora each of which represented one academic discipline. The Corpus Linguistics approach was applied to this study and the data were analyzed through using WordSmith tools (version 0.7). The quantitative results show that the abstracts in all disciplines show a similar word mean length, i.e. all of them is around (5). Qualitatively speaking, the results show that each discipline has its list of lexical words that are suitable for each discipline’s genre. The results also reveal a small amount of variation in terms of the tense of the reporting verbs specifically those which are used in the introductory part of the abstracts. However, the reporting verbs used in the body and the concluding parts of all abstracts are characterized by the past tense, third person, and active voice.

Keywords: Abstracts, Arabic, Corpus, Social sciences, Wordsmith

1. Introduction
Research articles have been recognized to be the most important constituent of scientific discourse where abstracts are acknowledged to be the important parts of research articles, and they are the focus of the author's ideas (Yang and Tian, 2015). Besides the title, the abstract is the first thing readers look for in research articles. Graetz (1982:p. 23) suggests that the
purpose of an abstract is to "give the reader an exact and concise knowledge of the total content of the very much more lengthy original, a factual summary which is both an elaboration of the title and a condensation of the report". Meanwhile, Swales (1990) indicates that the abstract is considered both a summary and a ‘purified’ reflection of the whole article. Bhatia (1993), on the other hand, points out that the informative function of abstracts is to present “a faithful and accurate summary, which is representative of the whole article” (p.82).

In addition to being informative, abstracts have a significant role in encouraging research reports. For instance, Holtz (2011) states that despite the typical aim for abstracts to summarize an article, they have become essential for readers in their decision process to read the text further, especially the case nowadays due to the vast amount of scientific publications. Furthermore, Hyland (2008) maintains that the abstract presents a decision-making stage for readers to determine whether the whole article deserves further attention or not. Therefore, Ren and Li (2011:p.163) demonstrates that many academics attempt to “persuade their readers to read the whole article by their effective selection of rhetorical features”. Martín-Martín (2005:5) also notes that “abstracts constitute, after the paper’s title, the reader’s first encounter with the text”.

On the other hand, some scholars have considered abstracts as having generic purposes in the research articles. For instance, Jordan (1991:p.508) views abstracts as “special narrow genre within the wider genre of description”. However, Lorés (2004: p.281) describes abstracts as "a genre in its own right which, while sharing many features of the research articles, also differs in several important aspects, one of which is its rhetorical structure”.

1.1 Aims of the Study

The current study aims to analyze quantitatively and quantitatively the Arabic journal articles’ abstracts written within the field of social sciences. It mainly intends to analyze the lexico-grammatical features of the examined abstracts in five academic disciplines; Economy, Geography, Psychology, Sociology, and Law. The features examined were: shallow features which reflect general characteristics of the corpus such as the word mean length; lexical features including the most frequent lexical items in each discipline; and grammatical features such as verb’s tense, voice, and person.

1.2 Questions of the Study

The increasing importance of abstracts in the academic field has motivated the researcher to do the current study which aims to do both qualitative and quantitative analyses of the abstracts found in social sciences articles. Particularly speaking, those abstracts accompanied the research articles in the following five disciplines; Psychology, Sociology, Geography, Economy, and Law. The abstracts were analyzed utilizing the methods of corpus linguistics and the use of WordSmith software. Therefore, the study aimed mainly to answer the following research questions:

1. What are the most frequent words in each discipline’s abstracts?
2. What are the upmost frequent reporting verbs used in each discipline’s abstract?
3. What are the frequent grammatical properties (tense and voice) of the reporting verbs in each discipline?

4. What is the most frequent concordance list in each discipline?

5. Do abstracts show differences in terms of the sentence and word length across disciplines?

1.3 Corpus Linguistics

Computers permit linguists to compile and quantitatively analyze large amounts of real texts. Such a collection of texts is known as a corpus, which is defined by Sinclair (1991) as "a collection of naturally-occurring language texts, chosen to characterize a state or variety of a language" (p.171).

Corpus linguistics gives a quantitative method to investigate human language employing naturally occurring language data. It performs several linguistic analyses by collecting a corpus consisting of a large number of machine-readable texts as well as utilizing specialized software programs that investigate frequencies, clusters, collocates, and concordances (Hunston, 2002). Theoretically speaking, corpus linguistics is concerned with developing corpora that are balanced, representative, and systematic (Hall et al., 2011). This implies that linguistic data are compiled and organized according to specific criteria related to the observed language and its speakers. As for the advantages of corpus linguistics, corpora allow us too rapidly and reliably search for linguistic data that are very difficult to obtain manually. Leech (1992, p. 107) and Biber et al. (1998, p.4) postulate that the core characteristics of corpus-based linguistic analysis are:

- It is empirical, analyzing the actual patterns of use in natural texts;
- It utilizes a large and principled collection of natural texts, known as a “corpus”, as the basis for analysis;
- It makes extensive use of computers for analysis, using both automatic and interactive techniques;
- It depends on both quantitative and qualitative analytical techniques;
- It focuses on linguistic performance and not on linguistic competence;
- It concentrates on describing language instead of prescribing it.

Furthermore, corpus linguistics has been closely correlated with other linguistic fields such as semantics, syntax, and discourse analysis to approach the different aspects of human languages. It has also been involved in the development of language engineering, language teaching, machine translation, and applied linguistics (Hunston, 2002).
1.4 WordSmith Tool

WordSmith Tool is a software package primarily designed for linguists and especially the case, for work in the field of corpus linguistics. It consists of a collection of modules for searching patterns in a language to see how words behave in particular texts. One of its main characteristics is that it can handle 80 languages such as Arabic, English, Chinese, etc and it runs under Windows.

The tools have been used by Oxford University Press for their lexicographic work in preparing dictionaries, by language teachers and students, and by researchers investigating language patterns in different languages in many countries around the world. (www.lexically.net/wordsmith).

The core areas of the software package include three modules:

Concord which is used to create concordances, so all the hits from a search within a previously defined body text.

WordList which lists all the words or word forms that are included in the selected corpus and statistical data is different from the text corpus.

KeyWord which creates a list of all those words and word forms according to certain statistical criteria in the text corpus significantly occurs rarely or frequently.

Each of the modules offers a number of other features concerning the text corpus or text being analyzed. Thus, for example, collocation and dispersion plots are computed with a concordance search. Besides, there are some additional modules that are useful for the preparation, cleaning-up, and formatting of the text corpus. WordSmith Tools is - along with several other software products similar in nature - an internationally popular program for the work based on corpus-linguistic methodology (www.wikipedia.com).

2. Review of Related Work

The journal articles’ abstracts have been attracted to the interest of those who are specialists in academic writing. Their interest has been reflected in their attempts to do descriptive linguistic analyses. One of the pioneer linguistic analysis abstracts was performed by Graetz (1982), who conducted a study aiming at analyzing the linguistic prosperities of 87 abstracts from the disciplines of health sciences, social sciences, education, and humanities. Graetz aimed to improve teaching practices for students of English as a foreign language, so they can successfully extract structural information from abstracts. Graetz’ findings demonstrated that the language of abstracts is characterized as follows:

The abstract is characterized by the use of past tense, third person, passive, and non-use of negatives. It avoids subordinate clauses, uses phrases instead of clauses, words instead of phrases. It avoids abbreviation, jargon, symbols, and other language shortcuts which might lead to confusion. It is written in tightly worded sentences, which avoid repetition, meaningless expressions, superlatives, adjectives, illustrations, preliminaries, descriptive details, examples, footnotes. In short, it eliminates the redundancy which the skilled reader
counts on finding in written language and which usually facilitates comprehension. (1982, p. 23).

On the other hand, Fluck (1988), analyzed qualitatively the linguistic features of abstracts of Economics, Linguistics, and Metal industry in German. The results showed that those abstracts were characterized by complex nominalizations, extensive noun compounding, and impersonality, use of the third person, passive voice, and present tense. Moreover, Jordan (1991) defined two types of abstracts; the descriptive and the informative ones. He aimed to provide linguistic criteria for the distinction between these two types. Nevertheless, this analysis was performed over a very small number of abstracts, and the criteria which were adapted for classifying the abstracts concentrated only on the use of passive voice and verb tenses.

Concerning the reporting verbs used in abstracts, Bloch (2010) examined the use of concordance to create material for teaching the role of reporting verbs in academic papers. He compiled two small corpora; one from the articles published in Science Journal and the other was from students in order to compare the performance of the students on the use of the reporting verbs with that of professional writers. The results of this study were then used as raw materials to design a database of sentences that could be used to create teaching materials for academic writing courses and to be available on the internet.

Furthermore, Wang and Tu (2014) investigated quantitatively and qualitatively the Journal Article abstracts in terms of tense use (i.e. verb variation) and rhetorical structure (i.e. move analysis). The sample of the study consisted of a corpus including 1,000 journal articles collected from four well-known journals in the field of applied linguistics; 250 abstracts from each journal. As for the verb use since it is the concern of the current study, the implemented data analysis showed that the most frequent type of verbs used in the abstracts was a verb (to be) followed by the reporting verbs; suggest, then show, examine, investigate, and finally find/found, respectively. Furthermore, the results revealed a variant but the consistent tendency of verb tense in each discipline. Verb’s tense varied from present, past, or passive voice.

Moving to Yang and Tian (2015) who examined evidentiality in 200 abstracts of English Research articles of four academic disciplines which are Philosophy, Computer, Linguistics, and Electronics. The study aimed to examine the lexico-grammatical properties of evidentiality in English abstracts to see whether there was an influential role of the academic discipline in the choice of evidentiality.

The results revealed that evidential was frequently common in English abstracts whereby writers use consciously various types to present their information and arguments. A comparison which was drawn among the four disciplines showed that the frequency of the use of evidentiality in abstracts of linguistics and philosophy was much higher than that in computers and electronics. Further, the analysis of reporting evidential and modal verbs in inferring evidential revealed that different disciplinary backgrounds of the writers have significant influences on their choice of evidentiality in their abstracts writing.
With reference to Portuguese Abstracts, the reporting verbs were examined by Feritas and Costa (2017). Their work was based on large Portuguese corpora which indirectly involved Computational Linguistics, as it was motivated by the quotation extraction task. In that, the verbs and patterns that were described in Feritas’and Costa’s study have already been integrated later in the development of an automatic quotation detection system for Portuguese. All the materials are public and are available for online consultations.

Based on corpora, the researchers established eight general patterns related to reporting verbs, from which it was possible to gather these verbs and then build a large lexicon of reporting verbs in Portuguese. This lexicon is composed of 308 verbs (the 293 verbs that were found in their research as well as the 15 verbs that were listed by Moura Neves but were confirmed by their study), and all verbs were manually validated through occurrences. According to the researchers’ points of view, their corpora provide a valuable contribution, which brings to light the potential of corpus-based descriptive studies on the Portuguese language.

As far as the researcher is aware, the brief review mentioned above showed an apparent absence of the Arabic studies regarding the linguistic analyses of Arabic abstracts in literature. Therefore, the current study may consider a valuable contribution to fill the gap in the literature on such topics and therefore may furnish grounds for further studies to analyze Arabic articles based on corpus linguistics. Further, the present study aimed to investigate five academic disciplines in the field of social sciences, unlike the previous studies which tackled a lesser number of disciplines.

3. Method

The study aims to characterize abstracts of social sciences’ articles through quantitative and qualitative analyses in terms of the lexical and grammatical features, using WordSmith (version 0.7) tools. The selected examined features were grouped into three categories; (1) shallow features which reflect general characteristics of the corpus such as the sentence length; (2) lexical features including the most frequent lexical items in each discipline; and (3) grammatical features such as verb’s tense voice and person.

The study built five sub-corpora where each of which represented one academic discipline. The five disciplines are; Economy, Geography, Psychology, Sociology, and Law. The corpus compromised of 500,000 words divided among the examined disciplines. The abstracts were collected from the following eight well-known academic journals: (1) Al-ʔafaq Liddiraasat Al-ʔiqtiṣadiyya, (2) Al-ʔbaqṣaad Al-ʔiqtiṣadiyya, (3) ʔattakamul Al-ʔ iqṣīṣadīyya, (4) Maḍzālat Al-ʔluuum Al-ʔidṭimaqṣīyya, (5)Maḍzālat Abḥaʔ Al-Yarmouk, (6)Maḍzālat Al-Buḥbūḥ Al-Juʔrafiyya, (7) Maḍzālat ʔilim ʔannafs/Maṣīr, (8) Maḍzālat Al-ʔluuum Al-Tarbaʔiyya, and (8) Maṇaṣat Al-Maḍzāllat Al-ʔilmīyya.

The corpora examined in the study were analyzed through using WordSmith tools (version 0.7) to find out the most frequent words appearing in each discipline, the statistical information e.g., the mean word length of each discipline, and the concordance list of the most frequent words especially those for the reporting verbs. Further, through using the
4. Results and Discussion

This section is designed to present and discuss the results revealed for each academic discipline in separate sub-sections. In each sub-section, the two WordSmith functions (Word List and Concord) were employed to the data provided for each discipline. The results are presented in the following sequence of disciplines; Economy, Geography, Psychology, Sociology, and finally Law. To illustrate, under each sub-corpora, the statistics list, the most frequent Wordlist, and the concordance list are presented through tables and figures. Furthermore, the list of the most frequent reporting verbs that appeared in each sub-corpora along with their concordance list are presented using figures.

4.1 Economy

4.1.1 Statistics List

Figure 1 shows the statistics list of the sub-corpora of the academic discipline Economy. It is noticed that the mean word length of the present corpora is ‘5.42’ and the word length std. dev is ‘2.30’.

![Figure 1. Statics list of economy sub-corpora](image)

4.1.2 Word List

The most frequent words found in the Economy sub-corpora are shown in figures 2 and 3. It is worth noting that having two figures of the frequency word list is due to the appearance of English words in which they were excluded.
Figure 2. Frequency word list in Economy sub-corpora

Figure 3. Frequency word list in economy sub-corpora

Table 1 presents the most frequent words that appeared in the Arabic Abstracts within the discipline of the Economy along with their English translations, number, and frequencies. It is worth noting that the function words and the words which are appeared in the meta-tags were excluded.

Table 1. Frequency word list in Economy sub-corpora

| Word             | English Translation | Frequency |
|------------------|---------------------|-----------|
| ?al?iqti?saadijja| Economical          | 916       |
| ?almaalijja      | Financial           | 451       |
| ?iqti?saadii     | Economic            | 414       |
| ?adduwal         | Countries           | 354       |
| ?attanmi?ja      | development         | 294       |
| ?almu?assaat     | Institutions        | 239       |
| Diraasah         | a study             | 286       |
Table 1 reveals obviously that previously mentioned frequent words are closely related to the Economy discipline itself. Especially those of ɁalɁistiθmaar (investment), Ɂattanmijja (development), and Ɂalmaalijja (financial). These words are also shown in the most frequent concordance lists with the word ɁalɁiqtiṣaadijja (economical). See the figures in the concordance section.

4.1.3 Concordance List

This section shows the concordance lists for the most frequent word appeared in the Economy discipline which is ɁalɁiqtiṣaadijja. See figures 4 and 5.

![Figure 4. Concordance list of ɁalɁiqtiṣaadijja (economical)](image1)

![Figure 5. Concordance list of ɁalɁiqtiṣaadijja (economical)](image2)
Figure 4 shows the appearance of the words: Ɂal Ɂazmah/Ɂal Ɂazamaat (crisis/crises) with the word Ɂal Ɂqtisadaa dija (economical). It is worth mentioning that this word usually collocates with the word economical in the Arab world because it is known that Arab countries suffer from certain financial difficulties and economic crises. Therefore, many conferences are annually held in Arab countries trying to provide solutions for these crises. One of the well-known conferences in the Arab World is ‘Davos’ which is held annually in Jordan.

4.1.4 Reporting Verbs

The most frequent reporting verb used in the introductory sentence of the abstract in this academic discipline is tahdif (aims) which has the frequency of (121) and it collocates with the word Ɂaddiraasah (the study). See figure 6.

![Concordance list of tahdif (aims)](image)

Figure 6. Concordance list of tahdif (aims):

On the other hand, regarding the reporting verbs used for presenting the results of the study, the results show that the most frequent verb is tawaasalat (came up with) and it collocates with the word Ɂaddiraasah (the study). See figure 7.

![Concordance list of tawaasalat (came up with)](image)

Figure 7. Concordance list of tawaasalat (came up with)
Further, the most frequent reporting verb used in the recommendation part of the abstract is ۡاوۡسات (recommended). Likewise ۡتاوۡسالت, the verb ۡاوۡسات collocates with the word ۡادیراسا ل (the study). See figure 8.

![Figure 8. Concordance list of ۡاوۡسات (recommended)](image)

To sum up, the results at hand reveal that in the field of Economy, the main linguistic features which characterize the reporting verbs used in the abstracts are as follows; the most frequent reporting verb appeared in the introduction part of the abstract is ۡتاهدیف and it is usually in the present tense, active voice and third person. On the contrary, the verbs which are used in both the results and recommendation parts are in the past tense, active voice, and third person.

4.2 Geography

4.2.1 Statics List

Figure 9 shows the statistics list of the sub-corpora of the academic discipline ‘Geography’. It shows that the mean word length is ‘5.33’ while the word std.dev. is ‘2.50’.

![Figure 9. Statics list of Geography sub-corpora](image)
4.2.2 Word List

Figures 10 and 11 show the lists of the most frequent words that appeared in the Geography sub-corpora.

Table 2 shows the list of the most frequent words in terms of number and percentages appeared in the Geography discipline along with their English translations. Both the function
words and the words that are related to a particular area such as (Al-Najaf City) were excluded.

Table 2. Frequency word list in geography sub-corpora

| Word             | English translations | Frequency | N  | %  |
|------------------|----------------------|-----------|----|----|
| Madiina City     |                      | 117       | 0.27|    |
| Diraasah a study |                      | 107       | 0.24|    |
| ؟almadiina City   |                      | 98        | 0.22|    |
| Manтиqah Area    |                      | 97        | 0.22|    |
| ؟almuhaafathah Governorate |    | 94        | 0.21|    |
| ؟almijjah Water   |                      | 88        | 0.20|    |
| ؟assukkan Population |                  | 79        | 0.18|    |
| ؟attabiicah Natural |                  | 70        | 0.16|    |
| ؟almunaaxijjah Climatic conditions |   | 69        | 0.16|    |
| ؟adduwal Countries |                  | 64        | 0.15|    |
| ؟alxassa؟is Characteristics | | 60        | 0.14|    |
| dardjah temperature |                 | 59        | 0.13|    |
| ؟dجuyraafii geographic |             | 58        | 0.13|    |

According to the definition of Geography, which is provided by the Wikipedia website, “is a field of science devoted to the study of the lands, the features, the inhabitants, and the phenomena of Earth”, the frequency of these words in this field is expected and clearly understood. It is worth mentioning that the word manтиqah (area) refers to the area which was examined in the research articles. Thus, manтиqah (area) frequently collocated with ؟اديرااسah. See the concordance section below.

4.2.3 Concordance

Figures 12 and 13 present the most frequent concordance lists for the most frequent words that appeared in this field which are; manтиqah (area) and ؟ادجعراافييjah (geographic). These words collocate with the words ؟اديرااسah (the study), and maةلعاامات (information), respectively.
4.2.4 Reporting Verbs

The results revealed that the most frequent reporting verb used in the introductory sentence of the abstracts in this academic discipline is tanawala (tackled) and it collocates with the words Ɂalbahθ (the research) and Ɂalbaaθθ (the researcher). See figure 14.
On the other hand, in the concluding part of the abstracts, the most frequent reporting verb used is *tawassalat* (feminine) or *tawassala* (masculine) which means (came up with) which collocates with *paddiraaasah* (the study). See figures 15 and 16.
As it is noted above, figures 14, 15, and 16 show that the reporting verbs which are appeared frequently in Geography abstracts are characterized by the past tense, active voice, and third person.

4.3 Psychology

4.3.1 Statics List

Figure 17 reveals the statistics list of the sub-corpora of the academic discipline Psychology. It shows that the mean word length is ‘‘5.14’’ while the word std.dev. is ‘‘2.24’’.

4.3.2 Word List

Figures 18 and 19 present the most frequent words which are appeared in the Psychology sub-corpora. The use of two figures to present the frequency word list is due to the appearance of English words that should be excluded from this study.

Figure 18. Frequent word list in psychology sub-corpora
Table 3 shows the list of the most frequent words along with their number and percentages which are appeared in the Psychology abstracts excluding the function words. Furthermore, the words Ɂannafs (psychology) and Ɂannafsii (psychological) are excluded because these words are the most frequent words that are something taken for granted.

Table 3. Frequent word list in Psychology sub-corpora

| Word             | English translation   | Frequency |
|------------------|-----------------------|-----------|
| Mistawa          | Level                 | 166       | 0.21     |
| Furuuq           | Differences           | 150       | 0.19     |
| Wujuud           | Existence             | 119       | 0.15     |
| Ɂalfard          | The individual        | 114       | 0.14     |
| Ɂalmuḍtamaq       | the society           | 112       | 0.14     |
| Ɂattifl           | Child                 | 106       | 0.13     |
| dʒamiçah          | University            | 104       | 0.13     |
| Dalaalah          | Indication            | 102       | 0.13     |
| Ɂaddirassat       | Studies               | 100       | 0.12     |
| Miqyyas           | Measure               | 100       | 0.12     |
| Ɂattarbijjah      | Education             | 98        | 0.12     |
| Ɂiḥṣaaʔiijjah     | Statistic             | 93        | 0.12     |
| Ɂalqalaq          | Anxiety               | 92        | 0.11     |
| Ɂatṭafkiir, Ɂaddakaaʔ, Ɂasuluuk | Thinking, Intelligence. Behavior | 80 | 0.10 |
The aforementioned list appeared in table (3) can be explained with respect to the fact that most of the research articles conducted in the field of psychology require the use of different types of measures to evaluate and assess cognitive concepts such as anxiety, thinking, intelligence, and behavior, etc. Therefore, the appearance of the words miqyas, Ɂalqalaq, Ɂattaʃkiir, Ɂaddəkaʔ, and Ɂassuluuk are expected. Further, for the study to be with a good value (i.e. significant), the results of psychological studies should show statistical differences, thus, words such as mistawa, furuq, wujuud are a must. The concordance lists mentioned in section 4.3.3 show further examples. The appearance of the words Ɂattifl (the child), fard (individual), and mudtamaq (society) are also expected because most of the studies in Psychology aim to investigate the children’s and societies’ behaviors. This is clearly shown in the definition of "Psychology" as:

the science of behavior and mind, including conscious and unconscious phenomena, as well as feeling and thought. It is an academic discipline of immense scope and diverse interests that, when taken together, seek an understanding of the emergent properties of brains, and all the variety of epiphenomena they manifest. As a social science, it aims to understand individuals and groups by establishing general principles and researching specific cases. (www.wikipedia.com).

4.3.3 Concordance

The following lists show the most frequent concordance lists for the most frequent words that appeared in the sub-corpora in the field of psychology. For example, Ɂannafsii (psychological) Ɂayyinah (a sample) and miqyas (measure). The frequency of Ɂayyinah (a sample) in the research articles of psychology is of great importance since as it is mentioned earlier most of the psychological studies aim to measure the cognitive behaviors of individuals and children and thus this cannot be done without a sample and a measure.

![Figure 20. Concordance list of Ɂannafsii (psychological)](image-url)
4.3.4 Reporting Verbs

The most frequent reporting verb used in the introductory sentence of the abstracts in the psychology academic discipline is hadafat (aimed) which has the frequency of (84) and it collocates with the word addiraasah (a study). See the figures below.
On the contrary, the most frequent reporting verbs used in the body, method, and the conclusion parts of the abstracts are Ɂasfarat (revealed), Ɂdar harat (showed), Ɂistaxdamat (used), Ɂudgriyyat (was conducted) and Ɂistanta Ɂat (concluded) respectively. It is worth noting that all of the former verbs are found to collocate with the word Ɂaddiraasah (a study). See the figures below.

Figure 23. Concordance list of hadafat (aimed)

Figure 24. Reporting verbs in the body and conclusion parts of the abstracts

Figure 24 shows that all the verbs used in the introduction, methodology, and conclusion parts of the abstracts are in the past tense, third person, and active voice except for one of the reporting verbs that is used to describe the methodological part of the abstract, i.e. the part describing the way a study is conducted is in the passive voice, e.g. Ɂudgriyyat (was conducted).
4.4 Sociology

4.4.1 Statistics List

The following statistics list for the abstracts in the field of Sociology shows that the word mean length is "5.27" and the word st. dev. is "2.28".

![Statistics list of sociology sub-corpora](image)

Figure 25. Statistics list of sociology sub-corpora

4.4.2 Word List

Figure 26 presents the most frequent words that appeared in the examined sub-corpora of Sociology. Further, table 4 reveals the list of the most frequent words appeared in this sub-corpora, excluding the function words and the expected words in this field such as ِالدَّيْمَاْيَنَّ / ِالدَّيْمَاْيَنَّ (social) or ِمَدْيَلاَْ (a journal) along with their number and percentages.

![Frequency word list of sociology sub-corpora](image)

Figure 26. Frequency word list of sociology sub-corpora
Table 4. Frequency word list of sociology sub-corpora

| Word                  | English translation | Frequency |
|-----------------------|---------------------|-----------|
| ئامعشيماça            | A society           | 231       |
| ئالهاجع                | life                | 128       |
| ئالپینسان             | human               | 117       |
| ئالکامل                | work                | 113       |
| ئالکعنف               | violence            | 113       |
| ئالکارابی              | Arabian             | 110       |
| ئالکیالام              | The World           | 106       |
| ئالدتیمااقی            | Sociology           | 99        |
| ئالپیسرع               | Family              | 97        |
| ئالدیووال               | countries           | 97        |
| ئالبکرییجع            | Human beings        | 94        |
| ئالمرگیاه              | woman               | 85        |

The most frequently used words shown in table 4 are highly recommended in the academic discipline of Sociology. As it is well-known that sociology is concerned with issues related to society including patterns of social relationships, social interaction, and culture (www.Wikipedia.com). Thus, these words are considered the heart of sociological research. For instance, ئالکعنف (violence) and ئالپیسرع (family) are considered the essential concepts of social interactions.

4.4.3 Concordance

Figures 27 and 28 present the concordance lists for two of the most frequent words that appeared in the examined sub-corpora which are ئامعشيماça (society) and ئالکعنف (violence).

Figure 27. Concordance list of ئامعشيماça (society)
According to what is appeared in figures 27 and 28, it can be noticed that the word ɁalmuʤātamaɁ (society) is frequently collocated with words such as Ɂamn (security), ɁinaaɁ (building), or qijjam (values). Whereas, ɁalɁunf (violence) is frequently collocated with its various forms such as ɁalɁunf Ɂid ɁalmarɁa (violence against women), ɁalɁunf Ɂazzawaadəiyi (Marital violence), and ɁalɁunf ɁalɁusarii (domestic violence).

4.4.4 Reporting Verbs

The most frequent reporting verbs used in the introductory sentence in the examined abstracts in the field of sociology is Ɂadafat (aimed) which is repeated (50) times in the sub-corpora. Ɂadafat is frequently collocated with Ɂaddiraasah (the study). See figure 29.
Meanwhile, the results revealed that the most frequent reporting verb used in the body part of the abstracts which concerns the presentation of the results is tawassalat (found) and it also collocates with the word Ɂaddiraasah (the study). See figure 30.

![Concordance list of tawassalat (found)](image)

**Figure 30.** Concordance list of tawassalat (found)

The analysis of the abstracts in the sub-corpora of sociology shows that most frequently used reporting verbs are characterized by the past tense third person and active voice.

### 4.5 Law

#### 4.5.1 Statistics List

The statistics list in figure 31 shows that the mean word length of the Law sub-corpora is ‘5.17’ and the word length std.dev. is ‘2.17’.

![Statistics list of Law sub-corpora](image)

**Figure 31.** Statistics list of Law sub-corpora
4.5.2 Word List

The following figures; 32 and 33 reveal the most frequent list of words that appeared in Law sub-corpora excluding the function words.

![Figure 32. Frequency word list of Law sub-corpora](image)

![Figure 33. Frequently word list of Law sub-corpora](image)

It is noted that most frequent words in this sub-corpora are considered the main parts of the field of law such as qanuun (Law), ɁalqaddaɁ (justice), ɁalmufarriɁ (legislator), ɁaldjinaaɁijjah (criminal), and Ɂalmaadah (an article in the law system). It is also shown that one of the most frequent words that are expected to appear is qaanuuniɁqaanuuniɁijjah (legal) and thus gets the high frequent list of concordance. See the concordance section below.
4.5.3 Concordance

The most frequent word qaanutunijjah (legal) and Ɂalmuʃarriğ (legislator) are also the most frequent words that get the frequent concordance lists. For example, qaanutunijjah collocates with different words such as Ɂalʔandimah (systems), Ɂalhimaajjah (protection), and qaawaqid (rules). See figures 32 and 33 below.

Figure 34. Concordance list of qaanutunijjah (legal)

Figure 35. Concordance list of qaanutunijjah (legal)
4.5.4 Reporting Verbs

The most frequent reporting verbs that appeared in the introductory part of the abstracts in Law sub-corpora is jatanaawal (tackles). This verb frequently collocates with the words addiraasah (study) or albahθ (research). See the concordance lists below.

However, the reporting verbs used in the body and the conclusion parts show a variation in the reporting verbs used such as xaradżat (came up with), Šxtatamat (concluded), Šntahat (ended with). See figure 38.
Accordingly, the abstracts which have been examined so far in the sub-corpora of law discipline show a great variation in the reporting verbs used in terms of tense. For example, the reporting verbs which are used in the introduction part are in the past form while those used in the body and conclusion parts are in the present tense. However, all verbs share the same voice and person i.e. active voice and third person, respectively. This may be attributed to the fact that the data collected for this discipline is smaller to some extent than the previous disciplines.

5. Conclusion

The Corpus Linguistics approach was applied to this study and the data were analyzed through using WordSmith tools (version 0.7). Each discipline was analyzed in a separate section where the WordSmith functions (WordList and Concordance) were applied to the examined data. By using these two functions, the statistical information regarding the data was provided and the concordance lists regarding the reporting verbs were provided and analyzed as well. The quantitative results show that the abstracts in all disciplines show a similar word mean length, for instance, all of them were around (5). Moreover, each discipline shows its list of frequent words, and this mainly because of the genre of the discipline itself which triggers the use of certain words. Furthermore, the abstracts reveal a small amount of variation in terms of the tense of the reporting verbs specifically those which are used in the introductory parts of the abstracts. For example, the most frequent reporting verbs used in Psychology, Sociology, and Geography disciplines are in the present form while those in Economy and Law are in the past tense. However, the results may indicate that all abstracts in the five disciplines show similarity in terms of tense that is used for the reporting verbs in the conclusion and body parts, i.e. past tense,. Moreover, all reporting verbs are characterized by active voice and third person.
References

Bhatia, V. K. (1993). *Analyzing genre: Language use in professional settings*. New York: Longman Publishing.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics. investigating language structure and use*. Cambridge: Cambridge University Press.

Bloch, J. (2010). A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research*, 2(2), 219-244. https://doi.org/10.17239/jowr-2010.02.02.7

Fluck, H.-R. (1988). Zur analyse und vermittlung der textsorte abstract. In C. Gnutzmann (Ed.), *Fachbezogener Fremdspracheunterricht* (pp. 67-90). Tubingen, Germany: Gunter Narr Verlag Tubingen.

Freitas, B., & Costa, S. (2017). Reporting verbs in Portuguese: A corpus-based descriptive study motivated by computational linguistics. *Fórum Linguístico*, 14(3), 2266-2285. https://doi.org/10.5007/1984-8412.2017v14n3p2266

Graetz, N. (1982). Teaching EFL students to extract structural information from the abstract. *International Symposium on Language for Special Purposes, August*(2-4), 1-23. Retrieved from http://www.eric.ed.gov:80/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/2f/f1/a9.pdf

Hall, C. J., Smith, P. H., & Wicaksono, R. (2011). *Mapping applied linguistics: A guide for students and practitioners*. London: Routledge.

Holtz, M. (2011). Lexico-Grammatical Properties of Abstracts and Research Articles. A Corpus-Based Study of Scientific Discourse From Multiple Disciplines. *Unpublished Doctoral Dissertation*. Darmstadt .

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. https://doi.org/10.1016/j.esp.2007.06.001

Jordan, M. P. (1991). The linguistic genre of abstracts. In A. Della Volpe (Ed.), *The 17th LACUS forum 1990* (pp. 507-527). Lake Bluff: IL: LACUS.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in Corpus Linguistic, Proceedings of Nobel Symposium* (p. 82). ACM, Berlin, New York: Mouton de Gruyter.

Lorés, R. (2004). On RA abstracts: from rhetorical structure to thematic organization. *English for Specific Purposes*, 23, 280-302. https://doi.org/10.1016/j.esp.2003.06.001
Martin-Martin, P. (2005). *The rhetoric of the abstract in English and Spanish scientific discourse: Across-cultural genre-analytic approach* (p. 279). European University Studies. Bern: Peter Lang.

Ren, H., & Li, Y. (2011). A Comparison study on the rhetorical moves of abstracts in published research articles and master’s foreign-language theses. *English Language Teaching, 4*(1), 162-166. https://doi.org/10.5539/elt.v4n1p162

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford University Press: Oxford.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings.* Cambridge: Cambridge University Press.

Wang, S. P., & Tu, P. N. (2014). Tense use and move analysis in journal article abstracts. *Taiwan Journal of TESOL, 11*(1), 3-29.

WordSmith Tools. (1996). *Windows software for finding word patterns.* Lexical Analysis Software & Oxford University Press.

Yang, L., & Tian, Y. (2015). A cross-disciplinary study of evidentiality in abstracts of English research article. *Open Journal of Modern Linguistics, 5*, 399-411. https://doi.org/10.4236/ojml.2015.54036.

Appendix A

**Phonetic chart for some Arabic sounds used in the study**

| Arabic Alphabet | Symbol | Description | Example | Meaning |
|-----------------|--------|-------------|---------|---------|
| ء               | ء       | voiceless glottal stop | ءامال | hope |
| ب               | b      | voiced bilabial stop   | باب   | door   |
| ت               | t      | voiceless dental stop  | تادج | crown  |
| ث               | θ      | voiceless interdental fricative | ثوىب | dress  |
| ج               | ð      | voice post-alveolar affricate | جمال | camel  |
| ح               | h      | voiceless pharyngeal fricative | حيىان | horse  |
| خ               | x      | voiceless velar fricative | خاتام | ring   |
| د               | d      | voice dental stop      | ديك   | rooster|
| ذ               | ð      | voice interdental fricative | ذوئب | wolf   |
| ر               | r      | voiced post alveolar liquid | رز   | rice   |
| ز               | z      | voice alveolar fricative | زيت   | oil    |
| س               | s      | voiceless alveolar fricative | سماى؟ | sky    |
| ش               | ð      | voiceless alveo-palatal fricative | شمس | sun    |
| ص               | ð      | voiceless emphatic alveolar fricative | صندوق | box   |
| ط               | ð      | voiceless emphatic dento-alveolar stop | طأوْلى | table |
| ظ               | ð      | voiced emphatic interdental fricative | ظيىل | shade  |
| Arabic | English |
|--------|---------|
| ١       | voiced pharyngeal fricative |
| ٢       | voiced uvular fricative |
| ٣       | voiceless labio-dental fricative |
| ٤       | voiceless uvular stop |
| ٥       | voiceless velar stop |
| ٦       | voiceless alveolar fricative |
| ٧       | voiceless alveolar nasal |
| ٨       | voiced alveolar liquid |
| ٩       | voiced velar stop |
| ١٠      | voiced palatal fricative |
| ١١      | voiced alveolar nasal |

**JSA consonant sounds with no responding SA orthography**

| Arabic | English |
|--------|---------|
| ١١      | voiced palatal fricative |
| ١٢      | voiced alveolar nasal |

**Vowels**

| Arabic | English |
|--------|---------|
| ١       | long low central unrounded |
| ٢       | short low central unrounded |
| ٣       | long high front unrounded |
| ٤       | short high front unrounded |
| ٥       | long high back rounded |
| ٦       | short high back rounded |
| ٧       | short high back rounded |
| ٨       | long high central unrounded |
| ٩       | short high central unrounded |

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/)