

A STUDY ON GENDER AND LANGUAGE DIFFERENCES IN ENGLISH AND ARABIC WRITTEN TEXTS

Hazim Alkrisheh

Multilingualism Doctoral School, University of Pannonia- Hungary

Feisal Aziez

Multilingualism Doctoral School, University of Pannonia- Hungary

Taisir Alkhrisheh

Amman Arab University- Jordan

ABSTRACT

This study aims at investigating gender differences in writing style. The study also aims at investigating language use differences in Arabic and English written texts by native speakers of Arabic in the average sentence length, lexical density, and readability. 40 students were asked to write an essay on the extent of effort expended to have better scores in academic settings. We used Halliday's framework about the functions of language to investigate gender differences. Halliday claimed that females' writing style is, as he described, 'involved' while males' writing style is more 'informative'. The results of the study do not confirm Halliday's assumptions about gender differences in writing. No significant differences were found between males and females in the frequencies of the use of nouns, prepositions, numerals and modifiers. The only significant difference that was found is the use of pronouns, which is not enough to account for the assumptions. To measure readability, the Gunning-Fog index formula was used. The results show that there was no significant difference between Arabic and English in the average sentence length, but there were significant differences in lexical density and readability. This result indicates that the Arabic written texts are lexically richer yet more comprehensible.

Keywords: *gender difference, language difference, written texts*

Sari

Penelitian ini bertujuan untuk menyelidiki perbedaan gender dalam gaya penulisan. Penelitian ini juga bertujuan untuk menyelidiki perbedaan penggunaan bahasa dalam teks tulis bahasa Arab dan Inggris oleh penutur asli bahasa Arab dalam hal rata-rata panjang kalimat, kepadatan leksikal, dan keterbacaan. 40 siswa diminta untuk menulis esai dalam upaya meningkatkan skor mereka dalam bidang akademis. Kami menggunakan kerangka kerja Halliday tentang fungsi bahasa untuk menyelidiki perbedaan gender. Halliday mengklaim bahwa gaya penulisan wanita, seperti yang dia jelaskan, lebih 'terlibat' sedangkan gaya menulis pria lebih 'informatif'. Hasil penelitian tidak mengkonfirmasi asumsi Halliday tersebut. Tidak ada perbedaan signifikan yang ditemukan antara pria dan wanita dalam frekuensi penggunaan kata benda, preposisi, angka dan pengubah. Satu-satunya perbedaan signifikan yang ditemukan adalah

penggunaan kata ganti, yang tidak cukup untuk mendukung asumsi Halliday. Untuk mengukur keterbacaan, rumus indeks Gunning-Fog digunakan. Hasil penelitian menunjukkan bahwa tidak ada perbedaan yang signifikan antara bahasa Arab dan bahasa Inggris dalam panjang kalimat rata-rata, tetapi ada perbedaan yang signifikan dalam kepadatan leksikal dan keterbacaan. Hasil ini menunjukkan bahwa teks yang ditulis dalam bahasa Arab lebih kaya secara leksikal namun tetap lebih dapat dipahami.

Kata kunci: *perbedaan gender, perbedaan bahasa, teks tertulis*

Received 05 March 2019 last revision 16 April 2019 published 31 May 2019
doi. 10.33603/rill.v2i2.2028

Introduction

Many researchers in the field of linguistics (namely psycholinguists and sociolinguists) believe that the females' choices regarding speech acts play an important role in the achievement of intimacy (Tannen, 1990). Females' choices regarding speech acts are identified as their way of maintaining relationships. In written texts, however, Halliday (1994) distinguishes between two types of differences among males and females. These two types are referred to as "involved" and "informative" writing. The former describes females as being involved in the sense that they assume that the reader knows the references in their written texts, and thus, the reader needs to be involved from the females' perspective. As a result, the reader senses a kind of personal and author involvement in the text. Males, on the other hand, tend to be informative in the sense that they provide more details about the things that are mentioned in the text because they assume that the reader needs background information, no matter how little, about the things being discussed in the text.

Languages are also different on many levels in syntax, phonology, phonetics, semantics and many other linguistic aspects. In written texts, to be specific, it is very important to keep track of the learners' ability to comprehend these texts in reading. Comprehension is defined as a cognitive process through which readers interact with the text to extract the meaning on the basis of their prior knowledge (Ruddell, 1994). That is why it is very important for teachers to provide the proper reading materials for their students, because, if the text is too easy or too difficult, the learner might lose interest. Readability is one of the aspects which will be used to investigate languages (by

languages in this paper we are referring to language use of both languages Arabic and English by native speakers of Arabic) in this paper. There are many formulas for detecting readability (see Dale & Chall, 2006a; Dale & Chall, 2006b; Flesch, 2006). The formula used to investigate readability in this paper is referred to as ‘Gunning-Fog index formula’ (Gunning, 1952) which is based on the average sentence length, the average syllable length and the average word length. The average sentence length will be investigated independently in addition to lexical density in this paper to either confirm or reject our hypothesis about Arabic and English.

In addition to investigating language use, this research also aims to investigate Halliday’s claims regarding gender differences in the two types mentioned earlier (involved and informative) in Arabic and English written texts. Many researchers examined the differences between males and females in controlled conditions. Controlled conditions provide suitable data in the sense that the researcher is provided with written text samples which can be described as similar especially when the participants are asked to comply to a certain number of words in the text (e.g. 400-500 words essays). There are also other objectives of this research that will be discussed later.

Writing style

A considerable number of studies focused on gender differences. Trudgill (1972) and Eckert (1989), for example, discuss lexical and phonological differences between males and females. Trudgill discusses the lexical choices of males and females based on a sociolinguistic variation of middle-class females and working-class males. He states “standard forms are introduced by middle-class women, non-standard forms by working-class men”. This statement suggests that the females’ choices of linguistic patterns indicate their tendency to use more prestigious speech acts. Whereas men tend to put little emphasis on prestigious speech acts as a result of inaccurate self-evaluation responses. Eckert, on the other hand, examined the phonological differences between males and females. He states “...sex is not directly related to linguistic behavior but reflects complex social practice”. The complex social practices that both genders display do not only affect linguistic choices, but they also affect behavioral choices.

Other researchers investigated pragmatic and phonological differences in informal writing styles and in speech acts displayed by males and females (see Holmes 1990; Key 1975; Labov 1990).

Many researchers in the field investigated language as a social phenomenon. Halliday is one of the pioneers in this domain as he introduces himself as a generalist who tried to look at language from each and every possible angle. Halliday's framework on the functions of language has been used for years by many researchers and linguists in the field. He introduces eight functions for language use (Halliday, 1975): 1) Instrumental, 2) Regulatory, 3) Interactional, 4) Personal, 5) Heuristic, 6) Imaginative, 7) Informative, 8) Divertive. According to Halliday, these functions are the main functions of language in all of its spoken and written forms. An example of Halliday's framework used in research is a study conducted on the basis of Halliday's register model (Lukin, Moore, Herke, Wegener & Wu, 2011). In this study, the researcher introduces the concept of "Register" as variation according to use from Halliday's point of view. The results of the study show that contextual settings constrain meaning potential. This study was based on spoken-based language data. But in written texts and written-based language, Halliday's framework focuses on other aspects regarding functions, namely "involved" and "informative" writing styles as functions in a written language.

Many studies were conducted to investigate gender differences in this framework. Parastoo Yazdani & Reza Ghafar Samar (2010) is an example of such studies. The study aimed at investigating differences between native and non-native male and female students from different universities in Iran. The study shows that non-native females significantly used more pronouns than non-native males. The results also show that there is no statistically significant difference between non-native males and non-native females in the use of specifiers. Native males and native females exhibited no significant differences in the use of pronouns or specifiers. The results also show that the female and the male stereotypical behavior is present as there were differences in the number of words, sentences' length and paragraphs' number between males and females. Even though the results of this study do not support Halliday's assumptions, yet they do not contradict these assumptions.

Ishikawa (2015) conducted a study to investigate gender differences among university students in argumentative essays. The students were asked to write an essay of 200-300 words in a controlled condition where the topic choice was restricted to two topics. The results of this study show that males used more nouns and thus, used more prepositions than females. The difference in the use of nouns and prepositions between males and females is statistically significant. The nouns used by males are associated to certain places, times and activities. The results also show that males used more numerals than females as demonstrated in previous research. Female students on the other hand, used particular personal pronouns more frequently than male students. They also used more modifiers (intensifiers and quantifiers) than males. This difference between males and females regarding pronouns was statistically significant. The words used by females are associated to psychological processes and feelings. The results of this study support Halliday's assumptions about "informative" and "involved" writing styles.

Readability

Turning to the literature on readability, many studies have been conducted using the different kinds of formulas for readability. One such research is conducted to investigate if the teacher's subjective judgment on the readability of texts she presents to her students matches the students' perspective on the readability of the text (Kako, 2018). The results of this study show that there is a high negative correlation between the teacher's subjective judgment and the students' perspective regarding the readability of the text. The teacher predicted (based on her own judgment) that the texts she presented to her students ranking from easiest to hardest would match the students' perspective on the level of difficulty. However, the findings show the opposite as students found that the texts that the teacher thought to be easy turned out to be hard and the texts that the teacher thought to be hard turned out to be easy from the students' perspective. The paper focuses on one measurement of readability known as 'cloze procedure' (Bormuth, 1967). This method uses a fixed numerated word in each sentence (e.g. 3rd word) in the text and the researcher deletes the numerated word in each sentence to see if the students can predict the words or not.

There are many formulas for detecting readability such as the Flesch–Kincaid (Flesch, (2006)) readability tests which is associated to the Flesch–Kincaid Grade Level. Most of these indices of readability rely on sentence length and word length. There is also the Coleman–Liau index readability test designed by Meri Coleman and Liau (1975) to detect the understandability of texts. This formula relies on the characters of the word rather than the syllables. The arguments against this formula is that character/syllable formulas are more accurate in detecting readability when done properly. The argument for this formula is that the measurement of characters is more accurate when the researcher uses a computer program than the measurement of characters and syllables. The Coleman–Liau index was designed to mechanically calculate samples of hard-copy text. Unlike syllable-based readability indices, it does not require an analysis of the character content of words, it only requires an analysis of their length in characters. As an advantage, it could be used hand in hand with theoretically simple mechanical scanners that would only need to recognize characters, words, and sentence boundaries, without the need for a full optical character recognition or a manual keypunching. Another formula designed to detect the understandability of the text is the Automated Readability Index (ARI) which produces a representation of the Grade level needed to comprehend the text as follows.

Table 1
Formula for detecting the understandability of the text (ARI)

Score	Age	Grade Level
1	5-6	Kindergarten
2	6-7	First/Second Grade
3	7-9	Third Grade
4	9-10	Fourth Grade
5	10-11	Fifth Grade
6	11-12	Sixth Grade
7	12-13	Seventh Grade
8	13-14	Eighth Grade
9	14-15	Ninth Grade
10	15-16	Tenth Grade
11	16-17	Eleventh Grade

12	17-18	Twelfth grade
13	18-24	College student
14	24+	Professor

This formula also relies on characters rather than syllables which makes it easier to be calculated with a computer program. The Gunning-Fog index (Gunning, 1952) is also another measurement for readability which also incorporates a grade level similar to the previous readability index ARI. This readability instrument is a good predictor of readability as it incorporates the number of complex words based on the syllable count. This is why this formula is adopted for measuring readability as it provides a balanced tool that considers the number of sentences, words and syllables. The last formula that will be discussed is the SMOG index for readability. Developed by McLaughlin, Harry (1969), this formula relies more on sentences' length and syllables' length. Some researchers in the field considered this formula better and more accurate than the Flesch-Kincaid formula as the later seems to underestimate the reading difficulty of a given text compared to the former (Fitzsimmons, Michael, Hulley, Scott, 2010).

Lexical density

Before conducting any form of investigation on lexical richness, it is important to define what is meant by 'word'. Read (2000); Nation (2001) identified the term 'word' in four different categories: a word family, a lemma, a type and a token, given in order from the most general to the most specific. A word family is a very broad term which refers to regular and irregular derivatives in any given language. In other words, a word family is a group of words that share a common base or root which incorporates the attachment of many prefixes and suffixes onto it (e.g. work, works, rework, worker, working, workshop, workmanship, etc.). A lemma associates the inflections to a base form. In other words, a lemma is a group of words that share grammatical associations (e.g. live, lives, lived, living). A type refers to the total number of unique words in a given text. A token refers to the total number of words in a given text. The difference between a type and a token is that a type considers only words without repetition in contrast to a token in which all words are considered even repeated ones.

According to Johansson (2009), lexical richness or type-token ratio (TTR) is best measured by considering the level of uniqueness in the vocabulary choices produced by speakers. This is one of the common measurements of lexical richness of the text and it refers to the ratio of distinctive and unique words to total number of words in a given text. The uniqueness of the speakers' words is a proper, but not an exclusive, predictor of proficiency. In a study conducted by Failasofah, & Dayij Alkhrisheh, (2018) to examine the lexical diversity and lexical sophistication of Indonesian students, a type-token ratio measurement is used to conduct the investigation by the aid of D_tools (Malvern, Richards, Chipere, and Duran, (2004)) to examine the lexical diversity, and P_Lex (Meara, and Bell, (2001)) to examine lexical sophistication. However, this paper aims to investigate language differences in lexical density in addition to readability and sentence length using online tools (available at: Mladen, A. 2006 text analyzer; and WebFX, 2018).

Halliday, 1985 defines lexical density as “the kind of complexity that is typical of written language”. Lexical density is highly associated to readability as it appears to be one of the factors that can determine the linguistic complexity of a written text. In other words, the less lexical density found in a text, the easier the text is to comprehend. Calculating lexical density is usually done by calculating word frequencies and category frequencies (see Laufer and Nation, 1995) in which the ratio of lexical items (words that have meanings) are considered against non-lexical items (e.g. articles such as ‘the’). For instance, in a sentence like ‘Mike loves going *to the* park’, the non-lexical items written in italics compromise the lexical density of the text and thus the lexical density of the sentence is 66.67%. Laufer and Nation (1995) had several measurements for measuring lexical richness such as Lexical Density (LD), Lexical Variation (LV), Lexical Originality (LO) and Lexical Sophistication (LS). Lexical originality, for instance, considers the number of unique tokens in the text. Lexical sophistication considers the number of advanced words in the text. Lexical variation considers the ratio of the number of different words to the number of repeated words. And finally, lexical density considers the percentage of lexical words in the text (i.e., nouns, verbs, adjectives, adverbs).

Research questions

1. Are there any significant differences between males' and females' writing style?
And, in case significant differences were found, do the findings support Halliday's assumption?
2. A- Are there any significant differences between Arabic and English in the average sentence length?
B- Are there any significant differences between Arabic and English in the lexical density?
C- Are there any significant differences between Arabic and English in readability?

Hypothesis

Based on the literature presented earlier. We hypothesize that the gender differences regarding 'involved' and 'informative' will be confirmed in this paper as we predict that female students will use more pronouns and modifiers than male students. Male students, however, will use more nouns (and prepositions) and numerals. Other objectives of the study include investigating the differences between Arabic and English in the aspects mentioned in the previous section.

Furthermore, we hypothesize that Arabic will have a less average sentence length than English. We also hypothesize that the lexical density in Arabic will be more than the lexical density in English. Regarding readability, we would hypothesize that the English text will be more comprehensible than the Arabic text given that English is not the mother tongue of the participants and thus they would use a simplified variety of English. These assumptions regarding Arabic and English are based on the fact that Arabic is a highly inflectional language. In other words, Arabic has many forms of derivations incorporating prefixes and suffixes more than other languages. For Example, the word 'أنلزمكموها' (pronounced: anulzimkumuha) in Arabic needs six words in English as an acceptable translation. The translation of the word in English is: 'Shall we bestow it upon you?'

Method

The current study is a cross-sectional descriptive study in which the participants' act (writing and essay) was controlled only by the number of the words. The design of the paper includes an observation (regarding writing habits) followed by a statistical analysis to examine the differences. The participants of this research were forty students (between 18 and 23 years old) from Mutah university in Jordan (N=40) whose native language is Arabic. The students were chosen on the basis of the objectives of this study which is to compare male students to female students in the writing style. The other objective is to compare Arabic to English in three aspects that can be identified in the average sentence length, lexical density and readability (using Gunning-Fog index formula for readability). On the basis of these objectives the selection of the students is as follows: 10 male Arabic language students, 10 female Arabic language students, 10 male English language students and finally 10 female English language students.

Procedure

The students were asked to write no more than three paragraphs of no more than 200 words on their efforts to achieve their educational goals (write about your acts of efforts for having better achievement results). In other words, they were specifically asked to write about the exertion of hard work that they display to achieve their educational goals. The exertion of hard work is displayed in the things you do to achieve your goals. The goals that are mostly discussed in the educational context usually refer to excelling in one's domain and getting high grades. The sentences that these students wrote fall under four categories. 1) Commitment. 2) Time management. 3) Mental and physical activities. 4) Irrelevant information. The first category, for instance, is displayed in sentences such as "I attend the class on time" which demonstrates the commitment to the time of the class. The second category is displayed in sentences such as "I study every day for 2 hours" which demonstrates the student's attempt to manage his schedule. The third category is displayed in sentences such as "I pay attention to the teacher" and "I sleep early so I can wake up early" which demonstrates a mental based or a physical based activity. The fourth category is displayed in sentences that do not fall under any of the previous three categories and has nothing to do with their main task of providing information on the exertion of hard work.

To investigate the differences between genders, an online text analyzer (Mladen, 2006 text analyzer) was used to calculate the frequencies of the words displayed in the text for male students and female students. Then the data of each student was imported to an independent excel sheet to calculate the frequencies using the 'SUM' formula. To locate the word categories, we used a different color for each category. Even though it is much easier to use the option 'Replace' in Microsoft excel for certain categories, we chose to highlight the words with different colors instead, because the option 'Replace' for the Arabic language is useless due to the fact mentioned earlier about Arabic being highly inflectional. For instance, prepositions and nouns appear independently in English, whereas in Arabic they can be conjoined in one word such as the word 'بِهَا' (pronounced: biha) which means 'about it' or 'about her'. Then, the formula was used to calculate the percentage of each category.

To investigate the differences between the two languages, another online text analyzer (WebFX, 2018) was also used in addition to the previously mentioned website (Mladen, 2006 text analyzer). The former site was used to check the readability (Gunning-Fog index formula) and the average sentence length of the text for both languages. We choose the Gunning-Fog index formula because it considers the sentence length, the word length, and the syllable length. The later website was used to check the lexical density for both languages (in addition to word frequencies for both genders). After all of these procedures, all data was imported to SPSS for conducting the analysis. 'Independent sample T-test' was used to investigate the differences between the Arabic and English, and between males and females.

Findings and discussion

The results in this section are presented in the following order:

- 1- Descriptive statistics for males and females presented in table 1 in M / F format.
- 2- The differences between males and females presented in chart 1.
- 3- Descriptive statistics for Arabic and English presented in table 2 in A / E format.
- 4- The differences between Arabic and English presented in chart 2.

The texts which the students provided were edited and corrected before making the analysis.

Table 2 Means and Std. Deviation

	Mean (Males / Females)	Std. Deviation (Males / Females)
Pronouns	10.92 / 17.09	5.89 / 5.86
Nouns	24.64 / 23.19	6.58 / 6.51
Prepositions	16.27 / 18.00	4.31 / 4.59
Numerals	1.40 / 1.18	1.97 / 1.46
Modifiers	9.17 / 11.42	3.23 / 3.91

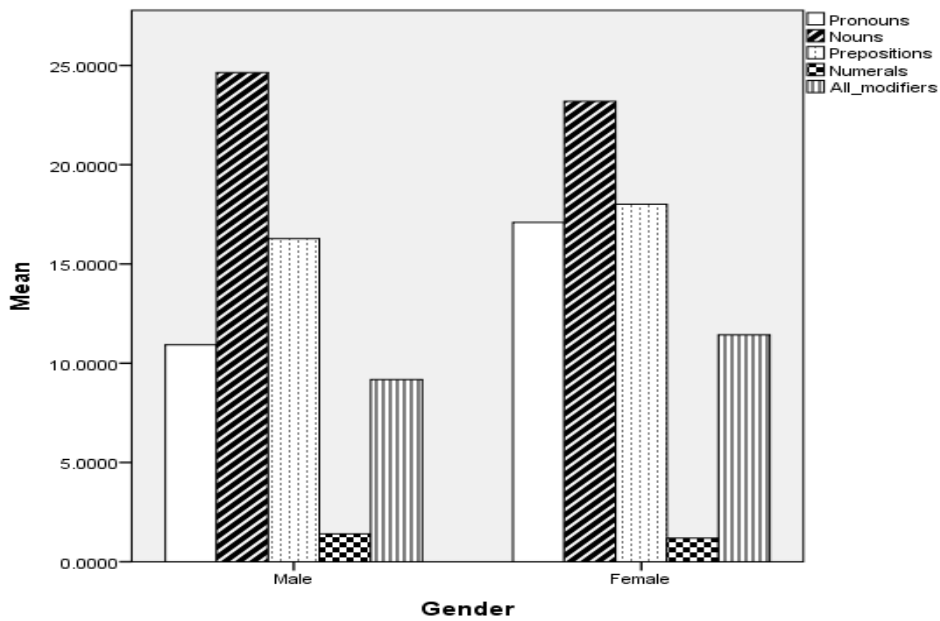


Chart 1 The difference male and female

Table 2

Table 3 Means and Std. Deviation

	Mean (Arabic / English)	Std. Deviation (Arabic / English)
Average sentence length	17.35 / 19.96	8.84 / 4.69
Lexical density	83.78 / 61.79	4.86 / 7.24
Readability	7.12 / 11.81	3.49 / 2.17

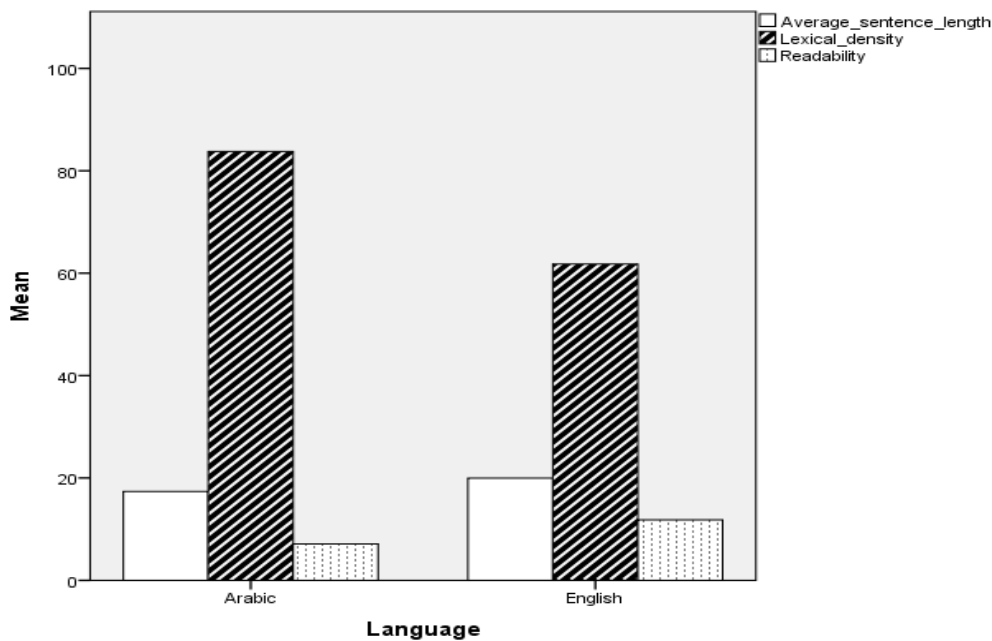


Chart 2 The difference of language

The following results are presented in relation to the research questions:

First, in table (1), the males' use of nouns is presented with a mean of (24.6) and a standard deviation of (6.5). The females' use of nouns is presented with a mean of (23.1) and a standard deviation of (6.5). Chart (1) shows that the difference between males (M=24.6, SD=6.5) and females (M=23.1, SD=6.5) in the use of nouns is not significant; $t(38) = .698, p = .489$. Chart (1) also shows that the difference between males (M=1.4, SD=1.9) and females (M=1.1, SD=1.4) in the use of numerals is not significant either; $t(38) = .387, p = .701$. Furthermore, chart (1) shows that the difference between males (M=16.2, SD=4.3) and females (M=18.0, SD=4.5) in the use of preposition is not significant; $t(38) = -1.227, p = .227$. However, chart (1) shows that the difference between males (M=10.9, SD=5.8) and females (M=17.0, SD=5.8) in the use of pronouns is significant; $t(38) = -3.315, p = .002$. And finally, chart (1) shows that the

difference between males ($M=9.1$, $SD=3.2$) and females ($M=11.4$, $SD=3.9$) in the use of modifiers is not significant; $t(38) = -1.978$, $p=.055$.

Second, to address the second question in its first subset, the results in table (2) and chart (2) show that the difference between Arabic ($M=17.3$, $SD=8.8$) and English ($M=19.9$, $SD=4.6$) in the average sentence length is not significant; $t(38) = -1.165$, $p=.251$. To address the second subset of the second question, the results in table (2) and chart (2) show that the difference between Arabic ($M=83.7$, $SD=4.8$) and English ($M=61.7$, $SD=7.2$) in lexical density is significant; $t(38) = 11.262$, $p=.000$. To address the third and last subset of the second question, the results in table (2) and chart (2) show that the difference between Arabic ($M=7.1$, $SD=3.4$) and English ($M=11.8$, $SD=2.1$) in readability is also significant; $t(38) = -5.096$, $p=.000$.

To address the first research question, it seems that Halliday's assumptions are not significantly confirmed in the results. Although males did use more nouns and numerals (but not more prepositions) but their use of these categories was not significant. According to Halliday's claims, males should use nouns, prepositions and numerals more than females. Even though it is true that males used more nouns and numerals than females, yet the frequency of use is not significantly established. There is one but not significant violation though, which is the use of preposition. Females used more prepositions than males. The argument about the use of prepositions is that the more nouns used, the more prepositions are likely to appear. So, it is more of a logical conclusion rather than an assumption about the use of prepositions. Turning to the use of pronouns and modifiers, the results regarding the use of pronouns indicate a significant difference between males and females, the results concerning modifiers on the other hand, is almost significant. The first research question is addressed with a negative answer since only one of the five categories account for the claims. These results do not significantly support Halliday's claims but they do not violate his assumptions. The reason behind this, as mentioned earlier in a previous section, is related to the number of the students as it has an impact on the results. Significant results could have been provided with larger numbers to represent the population.

Even though the average sentence length in the Arabic language scored less than the average sentence length in the English language as predicted in the hypothesis, yet the difference is insignificant. This result confirms our prediction in the hypothesis suggesting that Arabic is a highly inflectional language incorporating more prefixes and suffixes than other languages and providing more unique words deriving from the base form of the word such as 'ذهب' and 'ذهبت' pronounced 'ḍahaba' and 'ḍahabtu' respectively. The former means 'He went' or 'It went' and the later means 'I went'. The former has a null pronoun and the later has an attached pronoun as a suffix. This result suggests that Arabic is more readable than English. Even though Arabic scored more in lexical density, yet it scored less in readability, suggesting that the Arabic text contains more unique words, yet easily understood.

Conclusion

Regarding gender differences in this paper, it has been established in the results that the frequencies of the usage of the different categories is not significant. Even though Halliday's assumptions are not violated, yet they have not been significantly supported in this paper. There was a pattern in the use of word categories as suggested by Halliday though. The different roles that both genders play in society determine their linguistic and behavioral choices. The males' use of nouns and numerals, for instance, might be related to their need to confirm their authorial identity. The females' use of pronouns and modifiers might be related to their need to maintain intimacy and relationship. Regarding the differences between Arabic and English, it has been established in the results that there are significant differences in lexical density and readability. These results are an initial attempt to put Arabic in the field of linguistics and corpora because Arabic, as mentioned earlier, is an understudied language. More research should be conducted to confirm our hypothesis in a more fixated trend (e.g. controlled conditions for comparing languages). Research should also be conducted to investigate the differences between Arabic and other languages in all of its spoken and written based forms of native speakers. This initial attempt to present Arabic in its infancy in corpora is an important step, yet a step that requires more studies to confirm the assumptions. This research presents valuable data if supported with more studies and more research of an empirical nature. This research also represents a call for creating an Arabic corpus

by national and governmental institutions to support its existence in the field of linguistics even though it needs a methodological refinement due to its fresh presence in research.

Limitation

One identifiable downside to this research is the number of students who participated in the research. In any similar research, larger numbers are usually required to represent a better sampling of the population. Another downside that can be identified is the number of words required to write the texts. The reason behind this is that we didn't want to push the students over to provide irrelevant information, but some of them did. Some students provided unnecessary information such as 'teachers' strategies' used in class. The focus of the text was on the students' own efforts in pursuing their goals. Another downside to this research is the fact that the English language participants are not native speakers of English. For this reason, we decided to investigate written texts so that we can give the students enough time to express themselves properly as they tend to correct themselves all the time in written texts. This issue can be solved with future research on the differences between the Arabic and the English language of native speakers in spontaneous speech acts. The only reason for including non-native speakers is the availability of limited resources by which only the previously mentioned participants could be included. In other words, we do not have access to native speakers and so, we had to work with what we had. The selection of non-native speakers does not imply any form of bias and the results of this study can still be either confirmed or rejected with similar future research with native speakers of both languages

References

- Bormuth, J. R. (1967). Cloze Readability Procedure. *CSEIP Occasional Report (1)*.
- Coleman, Meri; and Liau, T. L. (1975); A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60, 283–284.
- Dale, E., & Chall, J. S. (2006a). A formula for predicting readability. In W. DuBay (Ed.), *The Classic Readability Studies* (pp. 63-74). Costa Mensa: Impact Information.

- Dale, E., & Chall, J. S. (2006b). A formula for predicting readability: Instructions. In W. DuBay (Ed.), *The Classic Readability Studies* (pp. 75-94). Costa Mensa: Impact Information.
- Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, 1(3), 245–268.
- Failasofah, F., & Dayij Alkhrisheh, H. T. (2018). Measuring indonesian students' lexical diversity and lexical sophistication. *Indonesian Research Journal in Education*, 2(2), 97-107.
- Fitzsimmons, P. Michael, B. Hulley, J. Scott, G. (2010). A readability assessment of online Parkinson's disease information. *J R Coll Physicians Edinb*, 40 (4): 292–6.
- Flesch, R. (2006). A new readability yardstick. In W. DuBay (Ed.), *The Classic Readability Studies*. Costa Mensa: Impact Information.
- Gunning, R. (1952). *The techniques of clear writing*. New York: McGraw-Hill.
- Halliday, M. A. K. (1994). *Introduction to functional grammar (2nd ed.)*. London: Arnold.
- Halliday, M. A. K. (1985). *Spoken and written language*. 1st ed. [Waurm Ponds], Vic: Deakin University.
- Halliday, M.A.K. (1975). *Learning how to mean*. London: Edward Arnold.
- Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language & Communication*, 10(3).
- Ishikawa, Y. (2015). Gender differences in vocabulary use in essay writing by university students. *Procedia - Social and Behavioral Sciences*, 192, 593-600.
- Johansson, V. (2009). *Lexical diversity and lexical density in speech and writing: A developmental perspective*. Working Papers in Linguistics.
- Kako, E. (2018). The Readability of EFL Texts: Teacher's and Students' Perspectives. Available at: http://www.academia.edu/37254404/The_Readability_of_EFL_Texts_Teacher_s_and_Students_Perspectives [Accessed 22 Oct. 2018].
- Key, M. R. (1975). *Male/female language*. Metuchen: Scarecrow Press.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2(2), 205-254.

- Laufer, B. and Nation P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Lukin, A., Moore, A. R., Herke, M., Wegener, R. & Wu, C. (2011). Halliday's model of register revisited and explored. *Linguistics and the Human Sciences*, 4(2), 187-213.
- Malvern, D., Richards, B., Chipere, N. and Duran, P. (2004). *Lexical diversity and language development: quantification and assessment*. Basingstoke: Palgrave Macmillan.
- McLaughlin, G. H. (1969). SMOG grading — a new readability formula. *Journal of Reading*, 12(8), 639–646.
- Meara, P.M. and Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect* 16 (3), 5-19.
- Mladen, A. (2006). *Text analyzer - text analysis tool - counts frequencies of words, characters, sentences and syllables*. Available at: <https://www.online-utility.org/text/analyzer.jsp> [Accessed 21 Oct. 2018].
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Ruddell, M. R. (1994). Vocabulary knowledge and comprehension: a comprehension-process view of complex literacy relationships. In R. B. Ruddell, M. R. Ruddell & H. Singer (Eds.), *Theoretical models and processes of reading*. Newark: International Reading Association.
- Tannen, D. (1990). *You just don't understand: Women and men in conversation*. New York, NY: William Morrow.
- Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, 1, 179–95.
- Uchida, A. (1992). When difference is dominance: A critique of the anti power-based cultural approach to gender differences. *Language in Society*, 21, 547-568.
- WebFX. (2018). *Free readability score test tool - readable*. [online] Webpagefx.com. Available at: <https://www.webpagefx.com/tools/read-able/readability-score.html> [Accessed 21 Oct. 2018].

Yazdani, P & Samar, R.G. (2010). Involved or informative: a gender perspective on using pronouns and specifiers in efl students' writing. *The Modern Journal of Applied Linguistics*, 2(5), 354-378.

Author's Biography

Hazim Alkhrisheh is a PhD student at Multilingualism Doctoral School, Faculty of Modern Philology and Social Sciences, University of Pannonia, Hungary. His research interests are language and motivation as well as corpus linguistics. He can be reached at hkhresha@yahoo.com

Feisal Aziez is a PhD student at Multilingualism Doctoral School, Faculty of Modern Philology and Social Sciences, University of Pannonia, Hungary. His research interests are second language development and TEFL. He can be reached at feiaziez@gmail.com

Taisir Alkhrisheh is a professor at Amman Arab University, Jordan. He can be reached at tkhrisheh@yahoo.com