

The Investigation and Comparison of “Orthographic Depth” in Persian and English Writing Systems through the Homography Feature

Abbas Ali Ahangar¹, Tahereh Taromi²

¹ *Corresponding author*, Professor, Department of English Language and Literature, Faculty of Literature and Humanities, University of Sistan and Baluchestan, Zahedan, Iran, Email: ahangar@english.usb.ac.ir

² Assistant Professor, Department of English Language Teaching, Farhangian University, Tehran, Iran, Email: t.taromi@cfu.ac.ir

Abstract

The concept of “orthographic depth” refers to the degree of the deviation of the writing system from the one-to-one correspondence between graphemes and phonemes. This study aims to explore and compare of the “orthographic depth” between Persian and English writing systems using UTPECC corpus and homographic scale. To this end, the scope of the application of the homography feature in 10000 words of Persian writing system was determined by using the word processing software 2010 through the synchronic study in the field of graphology. Then, the occurrence frequency of this feature was briefly compared with a similar number of words in English writing system. The research results show only 4 graphemes possess homography feature in Persian writing system, each of which corresponds to a limited number of phonemes (maximum 4 phonemes). While this feature is observed in 11 graphemes of English writing system and the variety of their corresponding phonemes was much more than Persian writing system. Considering the fact that the extent of the occurrence of the homography feature in English writing system is significantly higher than Persian writing system, it can be stated that the orthographic depth of English writing system exceeds that of Persian writing system.

Keywords: grapheme, homography, orthographic depth, phoneme, writing system

Received: April 12, 2023

Revised: October 28, 2023

Accepted: January 11, 2024

Article type: Research Article

DOI: 10.22111/ijals.2024.46415.2372

Publisher: University of Sistan and Baluchestan

© The Author(s).



How to cite: Ahangar A. A., Taromi, T. (2024). The investigation and comparison of “Orthographic Depth” in Persian and English writing systems through the homography feature. *Iranian Journal of Applied Language Studies*, 16(2), 129-144. <https://doi.org/10.22111/ijals.2024.46415.2372>

1. Introduction

The dominant approach in the study of writing systems is how to link language with written signs. In recent years, the writing system has gained special importance as a separate field in linguistic studies, and much research has been conducted in the field of various writing systems (Bijankhan & Alae Abouzar, 2013). However, the relationship between writing and language is not perfect and one-to-one, and almost all experts (including: Kearney, 1994; Rogers, 2005; & Safavi, 2016) believe that the correspondence between written signs and phonemes are relative because a script has a stable nature and is more conservative than language, and hence, over time, a gap is created between the spoken and written language and causes inadequacies in all writing systems. For instance; Natel Khanlari (1987, p. 277) states that “all of the common scripts of today’s world have deficiencies and maybe the script used for writing English has more flaws than the Persian script, but this does not cause English speakers to remain illiterate, nor does it prevent the spread of this language among other nations of the world”. The concept of “orthographic depth” refers to the relationship between language and writing (Rogers, 2005, p. 177), in other words; “orthographic depth” indicates the deviation of the writing system from the one-to-one correspondence between graphemes and phonemes, and the more difference between the written and spoken form of the language, the greater the orthographic depth (Rostami et al., 2012). The ratio of one to one (1:1) correspondence between grapheme and phoneme, which is considered its ideal state, is a state where a grapheme only indicates a corresponding phoneme and each phoneme is represented by only one grapheme, while in every writing system, there are various relationships between graphemes and phonemes, which causes the one-to-one correspondence between graphemes and phonemes to be disrupted, and as a result, increases the depth of the orthography. Among these relationships, the feature of “homography” is the concept that a single grapheme appears in different positions in the form of different single phonemes, which is one of the factors involved in the depth of the orthography. In this article, the authors seek to find the homography relationship by investigating the grapheme-phoneme correspondence in the data taken from the Persian-English comparative corpus data set of the University of Tehran (UTPECC) and based on the theoretical framework of Gontijo et al. (2003). Consequently, the scope of homography in Persian and English writing systems is determined and then the frequency of this feature in Persian and English writing systems is compared by using SPSS software. Hence, the hypotheses proposed in the present study are as follows:

- 1- The feature of homography is found in Persian writing system.
- 2- The feature of homography is found in English writing system.
- 3- There is a significant difference between the frequency of the homography feature in Persian and English writing systems.

This article consists of 6 sections. Section one was dedicated to the introduction. The second section introduces the related Iranian and non-Iranian literature review of the research. In section

three, the theoretical concepts and model of the research are given. Section four deals with the research methodology. In section five, Persian and English linguistic data under study are described and analyzed. Section six presents the discussion and conclusion.

2. Literature Review

Writing system is one of the issues highly rooted in culture and identity of a nation, and because of this, in most of the works dedicated to language and its history, writing system and its relationship with language have been mentioned in some way. Among the observed works, there was no study to specifically analyze homography feature. However, there are some works that, often without mentioning the title of homography, have referred to the types of corresponding relationships between graphemes and phonemes as well as the problems resulting from non-correspondence of grapheme-phoneme relationship.

2.1. Iranian Researchers' Studies

Despite the fact that most of the experts in the field of writing system, such as Natel Khanlari (1987) and Mahmoudi Bakhtiari (2000) confirm the existence of deficiencies in all writing systems and consider it a common feature of all writing systems, there are also people like Kaboli (2005), Haghshenas (2013), and Bateni (2019) who consider these disadvantages such as the one-to-one mismatch between the graphemes and phonemes in Persian writing system is so much that they aim at changing Persian script or have suggested its modification. Modarresi (1992, p. 180) claims "the writing system of Persian language (as an Indo-European language) does not fully correspond to the phonetic system of this language because this system has originally been programmed for the Semitic languages, which have considerable differences with Persian. On the other hand, it does not reflect the evolutions that have occurred in spoken Persian" (Modarresi,). Neisari (1995) has dealt with the representation of phonemic features of Persian in the domain of contrasting with Persian script and based on such a contrast he has determined rules for the grammar of Persian script.

Mahmoudi Bakhtiari (2000) has also discussed the features of Persian script and coherent planning for this script. He maintains that "the script is the written representation of the language, in such a way that it can be considered as the established manifestation of human speech, and that is why the discussion about this important linguistic fact is in the field of a linguist's work". He also adds that although linguists have considered writing as a secondary form of language, according to some people, for instance, Hagen (1963), in language planning, the importance of writing due to its communication role for individuals who are separated from each other in terms of time and place is more prominent than speech. Hagen (1963) imagines the lack of one-to-one correspondence between script and language not to be considered as a "problem and inadequacy" but it is as a

“general tendency” of all scripts, so it may be supposed as a “universal feature” of all writing systems. Danaye Tous (2004) has studied the effects of transparency and opaqueness of Persian script on the reading skills of normal Persian speaking children and the ones with dyslexia. Asi (2006) mentions a list of positive and negative features of Persian script, including the existence of several graphemes for a single phoneme and the correspondence of several phonemes with a single grapheme as well as the continuous and discontinuous writing of Persian written symbols. He has only stated the defects and problems of Persian script. Bijankhan and Alae Abouzar (2013) have explained the distance between written forms and their corresponding phonetic forms in Persian language in the framework of the concept of “orthographic depth”. In addition, by presenting the formula and mathematical method of segmentation of written pieces, they have tried to calculate the depth of Persian script.

2.2. Non-Iranian Researchers' Studies

Among non-Iranian researchers, there are many people who have addressed the writing system, script and the degree of correspondence between script and language, among which the following can be mentioned. Sampson (1985) while providing a precise definition of writing system in linguistics, points to the possibility of the existence of semantic writing systems. He believes “by adding more signs and conventions, a richer writing system can be made in a way that can express spoken language” (Sampson, 1985, p. 30). Coulmas (1989) considers all writing systems to have a certain complexity and mentions that the reason for this is the existence of fewer signs in the script compared to the phonemes of the language. Regarding the writing system of English language, he states that the relationship between graphemes and phonemes is more complicated than other writing systems and violates the one-to-one correspondence. A lot of research has also been done on grapheme-phoneme correspondence and interesting results have been obtained, among which Carney (1994) can be mentioned. By referring to the feature of homography, he shows that in English writing system, the grapheme <a> has this feature and can correspond to eleven different phonemes in different positions. Venezky (1999) points out that English writing system, like other writing systems, is facing some problems, and due to the complexities in it, efforts have always been made to reconstruct and simplify the script since the 13th century AD. Sproat (2000) has reconstructed the underlying representation of words by considering the phonological representation related to standard American English pronunciation and based on the generative phonological theory of the SPE model of Chomsky and Halle (1968). He has provided a method to calculate the orthographic depth of the English script and due to the significant difference between deep and superficial expressions in English language, he considers this writing system as a deep one. Fisher (2001) believes that all revised writing schemes such as Pitman, Bernard Shaw and Webster's alphabet have been created with the aim of creating more compatibility between the written and spoken forms of writing. Gontijo et al. (2003) is another research dedicated to

grapheme-phoneme correspondence in English writing system. In this article, with a computer analysis, the frequency of the occurrence of the corresponding phonemes of each grapheme has been determined. It is tried to predict the possible pronunciations for the grapheme chain by using these frequencies. Rogers (2005) provides explanations about the terms related to the writing system after stating the importance of writing in linguistics. While presenting examples of homography feature in the English writing system, he emphasizes that there is no writing system to have a perfect correspondence between its written and spoken signs. He also holds that the goal of the most of the plans proposed for the reconstruction of the English script is to create a one-to-one relationship between grapheme and phoneme, so that each phoneme is written by only one grapheme.

3. Theoretical Concepts and Model of the Research

In this section, some concepts in the field of graphology such as graphemes, relations between graphemes and phonemes, writing system and orthographic depth are pointed out, in addition, the theoretical model of the research is stated as well.

3.1. Grapheme

A grapheme is an abstract unit of the writing system that stands in opposition to other units of the writing system and indicates a unit of language. In other words, the grapheme is the smallest written unit that differentiates between a written minimal pair (Rogers, 2005, p. 10). While the term of “letter” refers to a set of graphemes that are used to represent consonants and vowels and are traditionally called “alphabet”. Different definitions and classifications for grapheme have been presented in works such as Coulmas (1989), Venezky (1999), and Safavi (2016), among which Venezky’s (1999) definition of grapheme is the basis of the current research. Because in this view, each grapheme corresponds to a phoneme, which is more suitable for analyzing the data of this study than the definitions given by other researchers.

3.2. The Relationship between Graphemes and Phonemes

It seems that at the beginning of the emergence of the alphabetic writing system, it was assumed that a simple grapheme should correspond to a single phoneme, but over time, this general principle has changed, so that today this ideal state of the alphabetic writing system is rarely found, and various relationships between graphemes and phonemes have emerged, and Persian writing system is no exception to this principle. One of the types of relationships between graphemes and phonemes is as follows: A single grapheme in different positions represents different phonemes. Such a relationship is called homography or heterophony feature (Safavi, 2016).

3.3. Writing System

In fact, the meaning of the writing system is a set of writing units and their usage conventions for writing a language (Sampson, 1985, p. 21). Therefore, the writing system of a language includes a set of signs (graphemes), a set of syntagmatic rules for signs and a set of rules for reading those signs. Thus, it can be seen that the most vital parts of a writing system are the script and its reading rules. The script includes a series of signs that are indirectly related to the language through phonetic signs (Haghshenas, 2013, p.14) and the reading system is a part of the writing system that comprises rules of reading the letters and determines how the letters should be read (Safavi, 2016, pp. 45-56). Every writing system must have a script and some spelling rules that make this system readable, nonetheless these two vital parts of writing system are not necessarily connected to each other. Therefore, different languages have the possibility of choosing similar scripts, although they apply their own writing system to it (Safavi, 2016, 64).

3.3.1. Persian Writing System

Iranian languages have been written in many different scripts during their long history, although various forms of Aramaic script have been predominant. Modern Persian is written in Arabic script, which is of Aramaic origin. In Persian writing system, the letters are written from right to left. It is different from a typical alphabet in that, only consonant sounds have a dedicated letter (Safavi, 2016).

Normally, Persian writing system is considered to consist of 32 letters. These letters represent consonants and long vowels of Persian language, while Persian short vowels usually do not have a sign in the script and are denoted with diacritics. (Neisari, 1995). In all languages, including Persian language, phonemes can be divided into two main classes, the consonants and the vowels. In the phonological system of the Persian language, there are 23 consonants and 6 vowels (Samareh, 2015, p. 81)¹.

/p/	/b/	/t/	/d/	/k/	/g/	/q/	/ʔ/	/s/	/z/	/š/	/ž/
/f/	/v/	/x/	/h/	/č/	/j/	/r/	/m/	/n/	/l/	/y/	
/i/	/e/	/a/	/u/	/o/	/â/						

As far as the difference between the number of letters/graphemes and phonemes of Persian language is concerned, it is clear that there is no one-to-one relationship between the letters/characters and phonemes of the Persian language. Some of the mismatch between the written and spoken forms of writing and the factors of increasing the orthographic depth are related to the feature of homography (one grapheme–multi phonemes).

¹ Whereas the phonetic symbols used in the present research are the ones used in IPA, the phonetic symbols of other Persian and English references presented in this article are the original phonetic symbols used in the relevant references.

3.3.2. English Writing System

English writing system is the alphabetic spelling system used by the English language. 26 letters of English script (originated from Roman alphabet) are written from left to right.

There are 24 consonants and 20 vowels (12 simple vowels and 8 diphthongs) in the contemporary English (British) phonetic system (Ahangar & Mojahedi Rezaeian, 2023, pp. 86-94), as given below:

/p/ /b/ /t/ /d/ /k/ /g/ /f/ /v/ /θ/ /ð/ /s/ /z/ /ʃ/ /ʒ/
 /h/ /tʃ/ /dʒ/ /m/ /n/ /ŋ/ /l/ /r/ /w/ /j/
 /i:/ /ɪ/ /e/ /æ/ /ə/ /ɜ:/ /ʌ/ /u:/ /ʊ/ /ɔ:/ /ɒ/ /ɑ:/ /ɪə/ /eə/
 /ʊə/ /eɪ/ /aɪ/ /ɔɪ/ /əʊ/ /aʊ/

44 phonemes of the English language are written by 26 letters. This difference indicates the lack of correspondence between English phonemes and script signs.

In addition, as to the significant differences between phonetic signs and written signs in English, the English writing system can be considered as a complex system like the Japanese, Akkadian and Egyptian writing systems (Coulmas, 2003, p. 183).

3.4. Orthographic Depth

The term “orthographic depth” refers to the distance of a writing system from a phonographic writing system (Bijankhan & Alaei Abouzar, 2013, p. 3). Among the alphabetical scripts, there are degrees in terms of correspondence between graphemes and phonemes, and accordingly the depth of the orthography changes. In general, the writing system with a one-to-one correspondence between graphemes and its phonemes is called a transparent (shallow) script, and a script whose graphemes and phonemes do not correspond one-to-one is called opaque (deep) script. Based on the transparency or opacity of grapheme-phoneme correspondence, alphabetic scripts can be divided in a continuum; on one side of this continuum are scripts such as Indonesian, German and Spanish (shallow scripts) and on the other side are scripts such as English and Hebrew (deep scripts) (Widjaja & Winskel, 2004, p. 37). A script like Persian script that contains clear and dark words at the same time is located at the middle of the continuum (Raman, 2006, p.18).

3.5. The Theoretical Model of the Research

In order to carry out the present research, the article of Gontijo et al. (2003) was chosen as the theoretical framework of the research². In this work, a complete table of the corresponding

² Due to the large amount of analysis presented in Gontijo et al. (2003), its presentation has been discarded here; however, its method of analysis has been applied for the data of the present research.

phonemes of the graphemes of the British English writing system has been presented through the computer analysis of a large volume of British English data while determining the frequency of grapheme-phoneme correspondence.

4. Research Methodology

In order to examine the hypotheses of the present research, the first version of the Persian-English comparative corpus of Tehran University was used. Comparative corpora are documents in two different languages that cover similar topics (for example, published news of common events). In order to make this corpus, two different sets of news from January 2002 to December 2006 from two separate sources (BBC news agency and Hamshahri newspaper) have been used. The BBC news collection included 53,697 documents with an average length of 461 words and Hamshahri news contained 191,440 documents with an average length of 527 words. According to the analysis that was supposed to be done on the research data, the Excel program was preferred over other programs and all the data were recorded in the Excel environment and the corresponding phonemes of each grapheme were inserted. For the pronunciation of Persian words, the authors' knowledge and intuition was almost enough; nevertheless, in cases where there was doubt, "Farhang Sokhan" was used as a guide. Oxford dictionary (1996) and Cruttenden (2014) were cited to determine the pronunciation of English words. The noteworthy point is that the pronunciation of all kinds of words was recorded independently and out of context. In this respect, all phonemic and phonetic processes occurring in their pronunciation were neglected. As to Persian data, there was not much problem in matching the written form (grapheme) and the spoken form (phoneme), but in the case of the English data, due to the complexity of this language, on one hand, and the writers' not being native speakers of English, on the other hand, there were some problems that were solved by referring to Guntijo et al. (2003). In this way, by examining the grapheme-phoneme correspondence, the graphemes with homography characteristics and their corresponding phonemes were determined in the Persian and English writing systems. In order to determine their frequency, the data body was transferred to the word processor 2010 software. This software provides the possibility of counting the searched graphemes by determining the number of repetitions of each grapheme along with highlighting the desired grapheme in all the words of the text so, the frequency of the occurrence of the graphemes under study (with homography feature) were extracted and their significance level was checked through SPSS software and Chi-square test.

5. Data Description and Analysis

In this section, after a detailed examination of the data and the determination of the items taking the feature of homography, the frequency of this feature and the phonemes corresponding

to each grapheme in Persian language are fully described. Then, procedures are briefly applied regarding the writing system of English language. Finally, the homography characteristic of the writing system of Persian and English writing systems is compared and analyzed.

5.1. Description and Analysis of Persian Data

Examining the grapheme-phoneme correspondence of Persian data shows that four graphemes <ا>, <و>, <ه> and <ی> have the feature of homography in this language. The next step was to determine the frequency of each of these graphemes and their corresponding phonemes, which are shown separately in this subsection in the form of frequency distribution tables.

5.1.1. The grapheme <ا>

The grapheme <ا> in different positions may represent four different phonemes. For example, in the word "اسب: /æsb/: horse", this grapheme represents the phoneme /æ/. In the words "اجازه: /eɟɑ:ze/: permission", "اتاق: /otɑ:q/: room", / and "باد: /ba:d/: wind", the letter <ا> represents the phonemes /e/, /o/ and /ɑ:/, respectively. As it can be seen in the statistics of table (1), the frequency of the phonetic realization of the letter <ا> in the form of the phoneme /ɑ:/ has a much higher frequency than other phonemes, so that among the analyzed data, the letter <ا> has been phonetically manifested in the form of the phoneme /ɑ:/ in 83.67% of the cases and the phonemes /æ/, /e/ and /o/ are followed based on the frequency of occurrence.

Table 1

The Frequency of the Phonemes Corresponding to the Grapheme <ا>

The corresponding phonemes of the grapheme <ا>	Frequency of occurrence	Percentage
/æ/	537	11.76
/e/	128	2.80
/o/	81	1.77
/ɑ:/	3819	83.67
total	4565	100

5.1.2. The grapheme <و>

In words such as "جواب: /ɟæva:b/: answer" and "ورزش: /værzeʃ/: sport", this grapheme represents the consonant phoneme /v/, in the words "خود: /xod/: self, خورشید: /xorʃi:d/:sun" it manifests the vowel /o/ and in the words "شور: /ʃu:r/: salty, زود: /zu:d/: soon" it is realized as /u:/. The phonetic manifestation of the grapheme <و> in the form of phonemes /o/ and /u/ is possible only in the middle and final positions of the word. But, in the form of phoneme /v/, it is possible for this grapheme to appear at the beginning, middle and end of the relevant words. The highest frequency of the phonemes corresponding to the letter <و> is assigned to /v/, and with a slight difference, /u:/ is placed after it, and /o/ is in the last rank with a frequency of 378 cases.

Table 2*The Frequency of The Phonemes Corresponding to The Grapheme <و>*

The corresponding phonemes of the grapheme <و>	Frequency of occurrence	Percentage
/v/	1129	46.87
/o/	378	15.69
/u:/	902	37.44
total	2409	100

5.1.3. The Grapheme <ه>

This grapheme may represent the vowel phoneme /e/ in words such as “پرنده: /pærænde/: bird, مدرسه: /mædrese/: school “or the vowel phoneme /æ/ in the negation word” نه: /næ/:no”. Also, in all the beginning, middle and final positions of the word, this letter can represent the consonant /h/, such as the words “هوا: /hæva/: air, مهر: /mehr/: affection and دانشگاه: /da:nɛsgɑ:h/: university”. The phonetic representation of the letter <ه> has the highest frequency in the form of the phoneme /e/ and has the lowest frequency in the form of the phoneme /æ/.

Table 3*The Frequency of the Phonemes Corresponding to the Grapheme <ه>*

The corresponding phonemes of the grapheme <ه>	Frequency of occurrence	Percentage
/h/	1094	38.92
/e/	1648	58.63
/æ/	69	2.45
total	281	100

5.1.4. The Grapheme <ی>

The grapheme <ی> may represent the consonant phoneme /j/ in words such as “یار: /jɑ:r/: friend and یک: /jek/: one “or it can represent the vowels /i/ or /ɑ/ in the middle and final positions of the words like “شیر: /ʃi:r/: milk and آبی: /ɑ:bi/: blue “and” موسی: /mu|sa/: Moses and حتی: /hætɑ/: even”. According to the statistics of the table (4), in 72.76% of the words under study, the letter <ی> represents the phoneme /i:/ and in 25.29% of the cases, it represents the phoneme /j/ and only in 1.95% of the data, it phonetically manifests in the form of the phoneme /ɑ:/.

Table 4*The Frequency of the Phonemes Corresponding to the Grapheme <ی>*

The corresponding phonemes of the grapheme <ی>	Frequency of occurrence	Percentage
/j/	724	25.29
/i:/	2083	72.76
/ɑ:/	56	1.95
total	2863	100

5.1.5. The Frequency of Homography Feature in Persian Writing System

When examining the graphemes with homography feature among the data of the Persian writing system, it can be seen that the grapheme <ل> has the highest frequency and the grapheme <س> and <و> are in the next ranks. The grapheme <چ> has the least amount of frequency. The frequency of phonemes equivalent to the grapheme <چ> is somewhat similar, but in the graphemes <ل>, <و> and <س>, their corresponding phonemes have a very different frequency compared to each other.

Table 5

The Frequency of Graphemes with Homography Feature in Persian Writing System

The graphemes with homography feature	Frequency of occurrence	percentage
< ل >	4565	36.09
< و >	2409	19.05
< و >	2811	22.22
< س >	2863	22.64
total	12648	100

5.2. Description and Analysis of English Data

All the steps mentioned in the description and analysis section of the Persian data were also carried out for the English data taken from the BBC news agency in the Persian-English comparative corpus of UTPECC. Due to the large number of phonemes corresponding to some English graphemes, the statistical data of the English writing system are briefly stated (the corresponding phonemes and their instances in the table 6 have been quoted from Gontijo et al. (2003, 147-157). Among the English data, there are eleven graphemes that have the feature of homography, among which the graphemes <a> and <l> have the highest percentage of frequency with 16.3% and 15.79% respectively. The graphemes <u> and <g> have the lowest percentage of frequency with 3.47% and 2.73%, respectively.

Table 6*The Frequency of Graphemes with Homography Feature in English Writing System*

The graphemes with homography feature	The corresponding phonemes	Frequency of occurrence	Percentage
<c>	/k/, /s/, /ʃ/, /tʃ/ case, cent, depreciate, cello	584	3.63
<f>	/f/, /v/ from, of	564	3.50
<g>	/g/, /dʒ/, /ʒ/ again, large, prestige	445	2.76
<n>	/n/, /ŋ/, /ŋ/ not, drink, recent	1926	11.97
<s>	/s/, /z/, /ʒ/, /ʃ/ last, these, pleasure, sure	1908	11.85
<y>	/I/, /j/, /aɪ/, /ə/, /iː/ mysterious, beyond, try, analysis, lycee	879	5.46
<a>	/æ/, /ə /, /ɑː/, /eɪ/, /ɒ/, /ɔː /, /eɪ ³ , /I/, /eə/ had, about, after, taken, was, all, many, manager, various	2572	15.98
<e>	/e/, /ə/, /I/, /iː/, /j/, /əː/, /eə/, /eɪ/ /ɑː/ get, after, because, female, azaleas, her, parent, elite, clerk	2269	14.10
<i>	/I/, /aɪ/, /ə/, /j/, /iː/ still, child, possible, million, litres	2523	15.67
<o>	/ɒ/, /uː/, /əu/, /ə /, /ɔː/, /ɪ /, /I/, /ɒ/, /au/ body, into, most, London, story, covers, women, woman, hour	1856	11.53
<u>	/ʌ/, /juː/, /jʊ/, /ə /, /w/, /ʊ/, /uː/, /I/, /e/ but, united, argument, until, language, full, revolution, busy, burial	571	3.55
total		16097	100

6. Discussion and Conclusions

In the present research, as aforementioned, the presence of heterophonic graphemes or, in other words, the feature of homography was investigated by examining the grapheme-phoneme correspondence in 10,000 words of the Persian language. The findings confirmed the first hypothesis of the current research; the feature of homography is assigned to four graphemes <|>, <س>, <و> and <و> in Persian writing system. Among them, the grapheme <|> has the most

³ /ɛ/= /e/

phonetic diversity in a way that it corresponds to four different phonemes, and the graphemes <و>, <و> and <و> are in the next ranks which may be used equivalent to three different phonemes in different positions.

In order to prove the second hypothesis of this research, the corresponding graphemes and phonemes of 10,000 English words were explored as well. The results show that eleven English graphemes include homography feature in English writing system and each of them might be equivalent with a variety of phonemes. The grapheme <a> has the highest frequency and the grapheme <g> has the lowest frequency. In this way, the second hypothesis of this study is also confirmed.

The investigation of the third hypothesis of this research is based on the Chi-square statistical test, the results of which are shown in Tables 7 and 8.

Table 7

The Results of the Statistical Test Comparing the Homography Feature in Persian and English Writing Systems

	Frequency of occurrence	Percentage
Graphemes including homography feature in Persian writing system	12648	44.0
Graphemes including homography feature in English writing system	16097	56.0

Table 8

Chi-square Test Results of Homography Feature in Persian and English Writing Systems

values	method
413.832	Chi-square statistics
1	degree of freedom
0.001	significance level

Based on the Chi-square statistics with the value of 413.832 and the degree of freedom with the value of 1, the significance level for the variable of graphemes including the feature of homography shows the value of 0.001. Since the value of the significance level corresponding to the Chi-square statistics is less than 0.05, therefore, the third hypothesis of the present study is confirmed because there is a significant difference in the frequency of the graphemes with homography characteristics in Persian and English writing systems.

The findings of the present research indicate that in some cases, the feature of homography (the graphemes that represent more than one single phoneme) has been exaggerated in Persian writing system and its existence is theoretically considered a fundamental problem. What can be observed in practice is that, firstly, these graphemes do not have a wide scope of use compared to other graphemes. Secondly, any Persian speaker who has mastered Persian writing system knows very well that in many cases, according to the position of these graphemes in the word, the equivalent phoneme can be predicted. For example, if the grapheme <و> or <و> is at the

beginning of the word, it will definitely be phonetically manifested as [v] and [j], and it is impossible to pronounce these graphemes as [o, u:] and [i:, a:] respectively.

It is obvious that the present article has reached such results only in the framework of graphology, and examining its possible difficulties in the field of education is another topic that requires its own research. Another finding of this research is that in English writing system, despite its widespread and international use, the feature of homography is found to a greater extent than in Persian writing system; in a way that eleven graphemes in English writing system have this characteristic and each may correspond to many types of phonemes. Perhaps it can be said that the feature of homography cannot be considered a deficiency and a problem; rather, it is one of the common features that exist in all common writing systems, including English. Mahmoudi Bakhtiari (2000) states that the feature of homography is the general tendency of all writing systems including Persian and English scripts, also it is considered as one of the linguistic universals, and the findings of the current research is in line with this claim.

Based on the finding of the research, it can be claimed that English writing system has more depth of orthography than Persian writing system. Because the frequency of homography feature in English writing system is significantly higher than Persian writing system. Thus this feature is one of the factors involved in increasing the orthographic depth. Generally speaking, it can be said that the results of this research agree with those of Coulmas (1989) and Sproat (2000) because they believe that in English writing system, the relationship between graphemes and phonemes is more complex and, as a result, the orthographic depth is greater than other writing systems. In this respect, the current research also shows that the frequency of homography in English writing system is significantly more than Persian writing system.

It is worth mentioning that in some previous research, often seeing homography feature as one of the defects of Persian writing system, its effect on reading and writing skills has been investigated. Similarly, no studies have explored the feature of homography in Persian writing system quantitatively. In addition, whereas there are many cases of the quantitative investigation of grapheme-phoneme correspondence and the feature of homography in English writing system, a study to compare the homography feature between Persian and English writing systems was not found.

References

- Abol-ghasemi, M. (2017). *A History of the Persian Language*. Samt.
- Ahangar, A. A., & Mojahedi Rezaeian, S. (2023). *English phonetics: A practical guide*. University of Sistan and Baluchestan Press.
- Asi, M. (2006). Persian in Raya Sepehr. *Nameye Farhangestan*, 8(3), 59- 70.
- Batani, M. R., (2019). *Language and Thought*. Agah.
- Bijankhan, M., & Alae Abouzar, E. (2013). Orthographic depth of Persian. *Language Research*, 1(4), 1- 19.
- Carney, E. (1994). *A Survey of English Spelling*. Blackwell.
- Coulmas, F. (1989). *The Writing Systems of the World*. Blackwell.
- Coulmas, F. (2003). *Writing Systems: An Introduction to their Linguistic Analysis*. Cambridge University Press.
- Crowther, J. (1996). *Oxford Advanced Learner's Dictionary*. Oxford University Press.
- Cruttenden, A. (2014). *Gimson's Pronunciation of English*, 8th Edition. Routledge.
- Danaye Tous, M. (2004). *Transparency of Persian Script and its Impact on Dyslexic and Natural Persian Speaking Children* [PhD Dissertation, Tehran University]. Tehran.
- Fischer, S. R. (2001). *A History of Writing*. ReaKtion Books Ltd.
- Gontijo, P. F. D., Gontijo, I., & Shillcock, R. (2003). Grapheme- Phoneme Probabilities in British English. *Behavior Research Methods, Instruments and Computers*, 35(1), 136-157.
- Haghshenas, A. M., (2013). *Phonetics*. Agah.
- Kaboli, I. (2005). *Persian Orthography*. Baztab Negar.
- Mahmoudi Bakhtiari, B. (2000). Persian Orthography: appropriate or not?. *Persian Language and Literature Teaching*, 55, 63- 70.
- Modarresi, Y. (1992). Language Issues and Language Planning in Iran. *Farhang*, 13, 177- 206.
- Natel Khanlari, P. (1987). *Linguistics and Persian Language*. Tous.
- Neisari, S. (1995). *Persian Script Grammar: A Study of the Relationship between Persian Script and Persian Language*. Tab' va Nashr.
- Raman, I. (2006). Are Acquired Dyslexias and Dysgraphias Language-Specific or Universal?. *The Iranian Journal of Contemporary Psychology*, 1 (2).
- Rogers, H. (2005). *Writing Systems, A Linguistic Approach*. Blackwell.
- Rostami, A., Rezaei, H., & Rafiei, A. (2014). A study of orthographic depth effect on Persian spelling errors in Azari speaking students. *Journal of Teaching Persian to Speakers of Other Languages*, 2 (5), 75- 98.
- Safavi, K. (2016). *An Introduction to Writing Systems*. Pezhvak.
- Samareh, Y. (2015). *Persian Phonetics*. Iran University Press.
- Sampson, G. (1985). *Writing Systems: A Linguistic Introduction*. Hutchinson.
- Sproat, R. (2000). *A Computational Theory of Writing Systems*. Cambridge University Press.
- Venezky, R. L. (1999). *The American Way of Spelling*. Guildford Press.

Widjaja, V., & Winskel, H. (2004). *Phonological Awareness and Word Reading in a Transparent Orthography: Preliminary Findings on Indonesian*. Macqarie University, Sydney.

