

Quantifying and visualizing the ambiguities of the semi-consonantal Persian writing system, and its consequences

by

Kaveh Ashourinia

Submitted to OCAD University in partial fulfillment of the requirements for the degree of

Master of Design in Inclusive Design

Toronto, Ontario, Canada, December 2018

© Kaveh Ashourinia, 2019

This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 Canada license.

To see the license go to <https://creativecommons.org/licenses/by-nc/2.5/ca/>

or write to Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

COPYRIGHT NOTICE

This document is licensed under the Creative Commons Attribution-NonCommercial 2.5 Canada License.
<https://creativecommons.org/licenses/by-nc/2.5/ca/>

YOU ARE FREE TO:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

UNDER THE FOLLOWING CONDITIONS:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

Non-Commercial — You may not use the material for commercial purposes.

Notices: You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

DECLARATION

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions, as accepted by my examiners.

I authorise OCAD University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

I further authorise OCAD University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

ABSTRACT

Literacy is more critical today than it has been at any other time, is a fundamental requirement for socio-economic inclusion, and acts as a force for prosperity in most contemporary societies. The scripts of a writing system connect the reader to the message, therefore we can think of them as components of written language.

Written Persian is a non-Roman alphabetical script system with roots in the Phoenician alphabet. It is an adaptation of Arabic script, even though the Persian language is very different from Arabic. For this and other reasons, many shortcomings and inconsistencies exist in Persian script as a way communicating the Persian language, and the script has been criticized throughout history for this reason.

This paper measures and demonstrates one of the shortcomings of the Persian writing system in terms of communicating the Persian language: the absence of short vowels and the inconsistency of long vowels. It introduces an analytical tool that could, with modifications, be applied to other alphabetic systems.

In the future, this tool may help inform investigation of the question of whether the Persian writing system can be improved through adjustments or if transitioning to another script is required.

Keywords: writing system, Persian language, literacy, Short Vowels, Long Vowels, orthography

ACKNOWLEDGEMENTS

My gratitude goes out to my advisors, **Dr. Peter Coppin** and **Richard Hunt**, whose support, expertise and contributions guided me throughout the process of this research. Both have been a source of inspiration and motivation.

In addition, I want to thank **Dr. Jutta Treviranus** for encouraging me to join the Inclusive Design program, and for supporting me since the beginning of this journey.

I also want to thank **Allegra Robinson**, who patiently copy-edited my final document.

My appreciation goes out to **Arezo**, who has been my supporter and source of encouragement, and without whom I would not be able to complete this project.

TABLE OF CONTENTS

Preface.....	1
1. Introduction	6
1.1 Background.....	7
1.1.1 Functional Literacy.....	8
1.1.2 Scientific Literacy.....	9
1.1.3 Quantifying Literacy and Earnings.....	10
1.2 Scripts and Alphabets.....	10
1.3 The Status of the Persian Writing System.....	13
1.3.1 My Approach.....	17
1.4 Terms.....	19
1.4.1 The Academy of Persian Language and Literature.....	19
1.4.2 Consonant Cluster.....	19
1.4.3 Diacritics.....	20
1.4.4 Digraph.....	22
1.4.5 Diphthong.....	23
1.4.6 Glyph.....	24
1.4.7 Grapheme.....	24
1.4.8 International Phonetic Alphabet.....	25
1.4.9 Ligature.....	25
1.4.10 Monophthong.....	25
1.4.11 Phoneme.....	26
1.4.12 Reading Direction.....	26
1.5 Scope and Limitation.....	27
1.6 Aim of This study and the Analytical Approach.....	28
1.6.1 The Argument.....	29
1.6.2 Analytical Approach.....	30

2. Quantifying and Visualizing the Consequences of Having No Vowels at the Word and Sentence Level in English.....	32
2.1 Step 1: Word Level	33
2.1.1 Bat, bit or bt.....	33
2.1.2 Farmer frmr	37
2.2 Word patterns	38
2.2.1 Words with only one consonant.....	39
2.2.2 Words with two consonants	46
2.3 Sentence Level	49
2.4 Implications of having no vowels in English orthography	53
3. Quantifying and Visualizing the Consequences of Absence of Short Vowel Letters in the Persian Writing System	54
3.1 Similarities and differences between the Persian and English writing systems	54
3.2 The Persian alphabet and writing system.....	56
3.2.1 Diacritical marks.....	60
3.2.2 Cursive writing	61
3.2.3 Vowel sounds.....	65
3.3 Ambiguities caused by the semi-consonantal Persian alphabet	67
3.3.1 Words with three consonant graphemes	67
3.3.2 Words with four consonant graphemes	73
3.3.3 Sentence level	73
3.4 Implications of having no short vowels in Persian orthography	77
4. Complications Resulting from Vowel Combinations and Vowels Representing Both Vowels and Consonants (<i>Mater Lectionis</i>).....	78
4.1 <i>Alef</i> , a long vowel and a carrier	78
4.1.1 <i>Alef</i> as the long vowel /o:/.....	80
4.1.2 <i>Alef</i> with letter the <i>Vāv</i> as the long vowel /u/	81
4.1.3 <i>Alef</i> with the letter <i>Ye</i> as the long vowel /i:/	81
4.1.4 <i>Alef</i> standing alone as a placeholder for <i>Zabar</i> /æ/.....	81
4.1.5 <i>Alef</i> standing alone as a placeholder for <i>Zir</i> /e/	82
4.1.6 <i>Alef</i> standing alone as a placeholder for <i>Pish</i> /o/	82
4.2 <i>Mater lectionis</i> in Persian.....	83
4.2.1 <i>Vāv</i> and <i>Ye</i> at the beginning of a word.....	84
4.2.2 <i>Vāv</i> and <i>Ye</i> at the end of a word	85
4.2.3 <i>Vāv</i> and <i>Ye</i> in the middle of a word	86
4.2.4 Two exceptions	86

4.3 Vowel combinations and digraphs.....	87
4.4 Unrepresented and unrecognized phonemes or diphthongs	88
5. Discussion.....	89
6. Conclusion	93
Bibliography	95
Appendix a: the IPA chart	98

LIST OF TABLES

Table 1.1: Diacritics make a difference to the basic shape of a letter	20
Table 2.1: Number of possible combinations of the letters ‘b’ and ‘t’ using inserted vowels between the consonants	35
Table 2.2: Word patterns based on one consonant and up to two vowels	40
Table 2.3: Possible word patterns for a single consonant phoneme and vowel graphemes	41
Table 2.4: The CMUdict phoneme table	42
Table 2.5: Number of possible word combinations with the CMUdict vowel phonemes	43
Table 2.6: A single consonant letter in combinations of digraphs and single vowel letters	45
Table 2.7: Vowel digraphs of the English writing system	45
Table 2.8: Number of possible words based on any kind of vowel sound using a single consonant	46
Table 2.9: Possible word patterns using two consonant phonemes	47
Table 2.10: Number of possible word combinations based on two consonant sounds	47
Table 2.11: Number of possible word combinations based on three consonant sounds	50
Table 2.12: Number of possible word combinations based on four consonant sounds	51
Table 2.13: Number of word combinations and possibilities for the sentence based on individual words	52
Table 3.1: The Persian alphabet chart, contextual forms and pronunciations	58
Table 3.2: Letter Be in different contextual forms	62
Table 3.3: Three Persian words composed differently in terms of cursive writing	64
Table 3.4: Letters Gāf, Re and Dāl	68

Table 3.5: The three consonants in the گرد combination	69
Table 3.6: Possible letter combinations for the letters Gāf, Re and Dāl	70
Table 3.7: Meaningful words made with combination of Gāf, Re and Dāl and pronunciations	72
Table 3.8: Number of possibilities for each word and the total number	76
Table 4.1: Alef and its different contextual forms	79
Table 4.2: Four words showing the letter Alef in different roles	80
Table 4.3: The six possibilities of using Alef at the beginning of a word	82
Table 4.4: The letter Ye as a consonant and as a vowel	83
Table 4.5: The letter Vāv as a consonant and a vowel	84
Table 4.6: The letter Vāv and Ye at the beginning of a word	84
Table 4.7: The letters Vāv and Ye at the end of a word	85

LIST OF FIGURES

Figure 0.1: A map of languages and writing systems in the Middle East region	3
Figure 1.1: Evolution of the letter K	13
Figure 1.2: A sample page from the Persian Writing Guideline	22
Figure 1.3: Scope of the research	28
Figure 3.1: Comparing two writing systems, the Persian and the English	55
Figure 3.2: Letters Be and Te	60
Figure 3.3: An example of Persian writing	62
Figure 3.4: Highlighted words in two groups	63
Figure 3.5: The seven Persian letters that only join preceding letters	63
Figure 3.6: Short vowel Zebar in the word برگ	66

Preface

I was born in the early 1980s in Tehran, Iran. At that time, the introduction of new technologies, including digital platforms, was a major driver of economic growth and globalization. IBM coined the term “personal computer”, or PC, early in the '80s when it launched consumer computers, changing the future of computing dramatically. The growing popularity of personal computers influenced and transformed every facet of modern life and established our dependence on these machines. I experienced previously established platforms becoming displaced by contemporary ones.

Technology was not the only societal element that was undergoing transition at the time. There were many transitions, and most were not peaceful. There were waves of political altercations both regional and global. Many regimes rose and fell during this period. Changeover, accompanied by conflict, took place both inside and outside of my country's borders.

Iran was one of the countries in the Middle East that experienced a transition in government. After Iran's revolution, the political stance of the country's leaders contributed to a war with its neighbour, Iraq, that lasted almost eight years. However, in spite of the social, political and technological changes, we in Iran continued to write in the same language, even though many other aspects of everyday life and culture had changed, including the way we thought and spoke. This raises the question of whether the traditional script serves the current needs of those who use the Persian language.

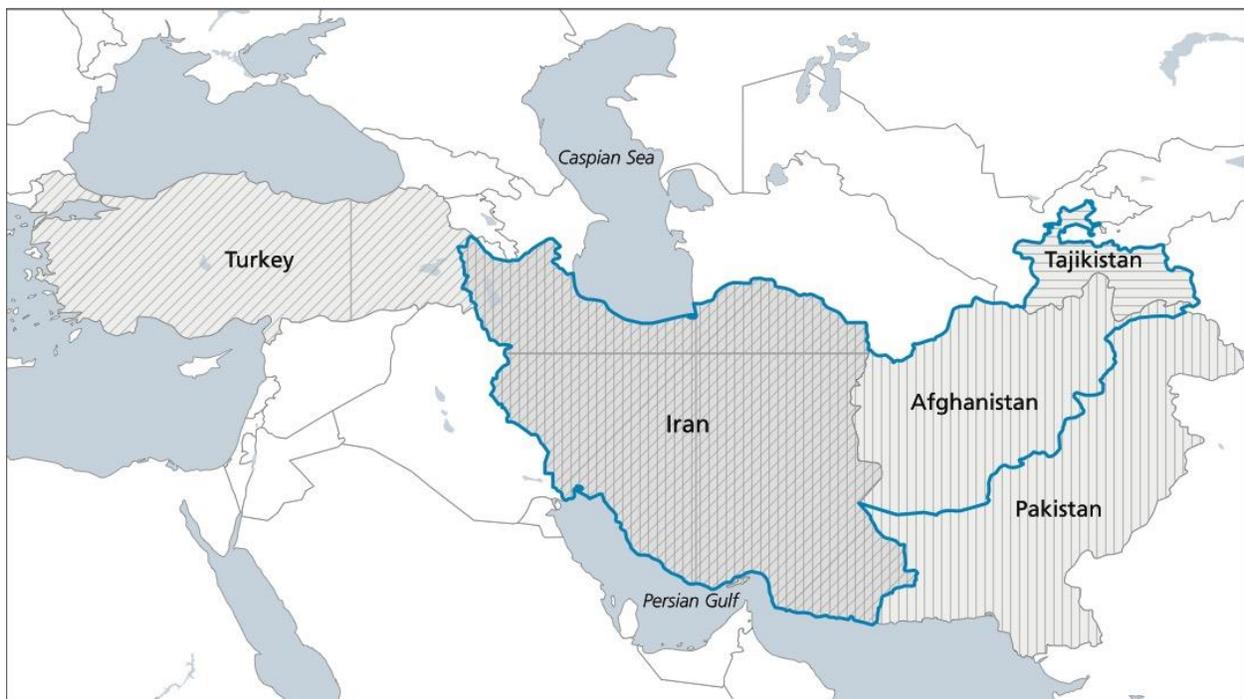
The language (both spoken and written) of the Iran that I grew up in is a hybrid of several influences. I want to explain this hybrid because my experiences with spoken and written language have influenced my interest in, and approach to, this project. The official spoken language of Iran is Persian, whereas the official writing system is Perso-Arabic (some people call it Arabic, while others call it Persian). The Perso-Arabic script is a direct descendant of Arabic script; therefore it has many characteristics in common with the Arabic alphabet. Although the official spoken language of Iran is Persian, Persian dialects, as well as other languages, are also spoken in Iran. These include Balochi, Arabic, Azerbaijani or Azeri, and some of the Kurdish languages. The geographic locations in which these languages and dialects are spoken are not divided by any official borders. However, all these languages and their variations share the same (official) Perso-Arabic writing system within the political borders of Iran.

I consider Persian to be my mother tongue or my first spoken language, but my mother's native language is not Persian. She was born in Tabriz, a city northwest of Iran, and her mother tongue is Azerbaijani, which is descended from neither Persian nor Arabic. In Iran, Azerbaijani is written in Perso-Arabic; but in Azerbaijan, it has been written in a modified Roman alphabet since the area became independent from the Soviet Union (Cyrillic was used before independence). Because my mother's family migrated to Tehran and she was educated there, my mother is more comfortable speaking Persian as a principal language and using Perso-Arabic as her principal writing system. I've always found it interesting that a writing system of one tradition can be used to convey a spoken language from another tradition. In this case, the Persian language, while quite different from the Arabic language, makes use of Arabic.

The Middle East region is full of different languages and cultures. There are languages that are very similar that use different script systems, while some languages that are very different from one another share the same, or a similar, writing system. In the paragraph above, I mentioned just a small sample of the languages and official writing systems of Iran and a few of its neighbours. The full situation of the region is, of course, more complex. As visualized in the following map, for example, Iran, Afghanistan and Tajikistan are labelled as Persian/Farsi-speaking countries (different branches of the

Persian language are spoken there). Tajikistan, unlike the other two countries, uses a modified version of Cyrillic for Tajiki Persian, while Pakistan uses Perso-Arabic scripts for the Urdu language. Another interesting example in the region is Turkey. Its official writing system was based on the Perso-Arabic script during the Ottoman Empire, but this changed to the Roman alphabet after Atatürk's reform, in the early 20th century.

This map shows a few of the many examples of similar situations. Of course, the same can be said of many other scripts. The Roman script is used for many languages, including some that are largely unrelated to Latin, notably, for Vietnamese, a language from the Austroasiatic family. Chinese has been used for other languages in Southeast Asia, and continues to be used in Japanese writing, and similarly Cyrillic deployed to be used for many languages.



- Branches of Persian language
 - Similar Writing System
- Different Writing System
 - Formerly Similar Writing System

Figure 0.1: A map showing some examples of languages and writing systems in the Middle East region (Map by Free Vector Maps)

I began my endless and no-return journey years ago. Since that time, I have been exposed to a variety of cultures, languages, ideologies and communication platforms. From my point of view, the latest part of my exploration is the most interesting. Two fascinating phenomena of our current era are the dependence of modern society on literacy, and the independence of writing itself from a specific language.

I think one of the best ways for different groups to avoid conflict and tension is to facilitate clear communication, since misunderstanding is the source of many problems. In this, writing, as both a carrier for spoken language and a platform for communication, plays an important role. At the same time, modern life dictates ever-higher levels of education and literacy for people who seek inclusion in any society. On a global scale, countries compete to gain knowledge and advance their scientific and intellectual agendas, which requires strong literacy. Preventing a society from learning foreign languages, or certain groups of foreign languages, is a tool that has been used by regimes to create a linguistic iron curtain and minimize the influence of other countries on a population. Literacy helps the flow of knowledge and scientific theories between academies, scientific institutions and industries. And a clear written language is a writing system with a minimum number of introduced errors, ambiguities or other problems.

I believe the Persian writing system has some serious problems that have led to its failure to answer the needs of modern life. These problems have a negative economic and social impact on people and their daily tasks. As a result, society as a whole suffers both nationally and globally, compared to advancements of some other societies. This will be explained further in Section 1.3: The status of the Persian writing system.

My final point is that, as I mentioned earlier, there are many languages spoken by large numbers of people in Iran, and there is a wide range of dialects of the Persian language both inside and outside Iran. I respect all these variations, and I believe in the beauty of this colourful and diverse culture. But due to the nature and scope of this research project, I am using Tehrani's dialect for referencing words and examples, unless

additional examples are required (in which case I will point this out). I chose to use the dialect of Tehran because I was born and grew up there, not because I value any cultural tradition over another.

1. Introduction

Literacy, or the ability to read and write, is a fundamental requirement for socio-economic inclusion and acts as a force for prosperity in most contemporary societies. A job that does not require literacy or a day without exposure to written documents is rare in everyday life, which shows how modern societies rely on and depend on literacy. Factors that promote or hinder literacy are therefore important in the success of a society. The aim of this project is to consider and quantify the degree to which Persian orthography, particularly the treatment of vowels, may interfere with clear understanding of the representation of words. Because the Persian writing system does not fully represent the complete sound of words, due to the fact that it does not represent short vowels, it is likely to make reading harder, particularly for children and others first learning Persian, and thus likely to discourage the acquisition of literacy.

Also, the needs of modern literacy are different than they have been in the past. Users of scripts are more likely to be engaged in short, conversational exchanges (most notably in texting and emails), where the existence of a context that often mitigates shortcomings in representation of phonemes is absent. Navigational typography, such as wayfinding and maps, where context is often absent, also makes the lack of correlation between the complete word and its representation in script more of a problem.

The rest of this chapter considers definitions, the background of the issue, and my approach to analyzing the degree to which certain characteristics of the Persian writing system fail to effectively represent the Persian language.

1.1 Background

Literacy is more critical today than it has been at any other time in the history of civilization. According to the United Nations Educational, Scientific and Cultural Organization (UNESCO), literacy is a human right. The first principle of UNESCO's Report of the International Commission on Education for the Twenty-first Century states:

education is a basic human right and a universal human value: learning and education are ends in themselves, to be aimed at by both individuals and societies and to be promoted and made available over the entire lifetime of each individual. (UNESCO, 1996)

In 1997 UNESCO replaced the word "basic" with "fundamental" in the final report of the Fifth International Conference on Adult Education, the Hamburg Declaration on Adult Learning, and emphasized the importance of literacy as a life skill.

Literacy, broadly conceived as the basic knowledge and skills needed by all in a rapidly changing world, is a fundamental human right. In every society literacy is a necessary skill in itself and one of the foundations of other life skills. ... Literacy is also a catalyst for participation in social, cultural, political and economic activities, and for learning throughout life. (UNESCO, 1997)

In addition to being a fundamental human right for individuals, literacy is an ability that is essential for the entire society to move forward and participate in economic and social progress on the global scale. As UNESCO's report "Mobile Phones & Literacy" states:

Literacy can encompass autonomous/neutral skills¹, functionally applied in context, yet can also be taken forward from a local/individual learning experience to include participation in wider economic, social, cultural and political spheres. Literacy is inherent to a continuum of learning in connection to society and a human right essential for lifelong learning and social change. (UNESCO, 2015)

The term literacy can be broken down in order to better evaluate the status of knowledge and the literacy rate of an individual or a society.

1.1.1 Functional Literacy

Merriam-Webster's definition of a functional illiterate is "a person who has had some schooling but does not meet a minimum standard of literacy" (Merriam-Webster.com, 2018). According to the research by the government of Canada, in 2012 only "51.5% of Canadians aged 16 to 65 had literacy scores in the Level 3 category or above" (well-being.esdc.gc.ca), which means that 48.5% of the population in that age group scored in at Level 2, 1 or below. Based on this measurement, literacy ability is divided into 4 levels.

The Canadian Encyclopedia defines these levels as (Bailey et al, 2013):

- In level 1, people have difficulty dealing with printed materials and have few basic skills for decoding or working with text. They are identified as people who cannot read.
- In level 2, people can use printed materials only for limited purposes such as finding a familiar word in a simple text that is clearly laid out. They are recognized as having difficulties with common reading materials.

¹ Autonomous or neutral skills are deemed to be independent of social and individual backgrounds and contexts and are expected to be acquired by individuals along standard cognitive processes of learning. These are the so-called basic literacy skills – reading and writing, and numeracy – as well as digital skills. (UNESCO, 2015)

- In level 3, people can use reading materials in a variety of situations if the tasks are not too complex.
- In level 4, people have strong literacy skills. This is a large and diverse group that shows a wide range of reading skills and many strategies for dealing with complex materials. These people can meet most reading demands and challenges.

While the main focus of this paper is on the Persian writing system, it is worth mentioning the literacy rate in Iran and a few of its neighbouring countries. According to The Institute for Statistics of UNESCO (UNESCO, 2013), and The World Bank (The World Bank, online database), in 2008 the literacy rate was 85% in Iran and 55% in Pakistan. In 2011, it was 32% in Afghanistan and 78.5% in Iraq. These four countries all share the Perso-Arabic script system.

Measuring a notion like functional literacy or scientific literacy in some of these places is not easy or even possible. But data show that the overall literacy rate in Canada is around 97% (Hammer, 2018), and considering the large gap between this number and literacy level 3 and level 4 in Canada, we can assume that similar proportions of literacy would be applicable in those places.

1.1.2 Scientific Literacy

Another term for literacy is scientific literacy. Many aspects of this term are tied to our daily contemporary lives.

Scientific literacy means that a person can ask, find, or determine answers to questions derived from curiosity about everyday experiences. It means that a person has the ability to describe, explain, and predict natural phenomena. Scientific literacy entails being able to read with understanding articles about science in the popular press and to engage in social conversation about the validity of the conclusions. (p.22, National Science Education Standards)

Speaking of scientific literacy, it should be pointed out that the reading direction of mathematical notation and scientific convention is left-to-right, in contrast to the right-to-left reading direction of Persian scripts. This particular area of writing may cause its own difficulties and merit investigations that are not part of this paper.

1.1.3 Quantifying Literacy and Earnings

Although quantifying the direct impact of literacy on quality of life and average income is not an easy task three economists at the University of Padua, Giorgio Brunello, Guglielmo Weber and Christoph Weiss, undertook it. Their study on adult men suggests that people who grow up surrounded by books are likely to earn more money. The study also showed that each additional year of education “increased a man’s average lifetime earnings by 9%. But the returns varied markedly according to socio-economic background”. Jamie Doward writes that “men brought up in households with less than a shelf of books earned only 5% more as a result of the extra year’s education, compared with 21% more for those who had access to a lot of books” (Doward, *The Guardian*, 2016). This study emphasizes how the role of early education has a direct relation to literacy.

Obviously, it would be naïve to say that literacy, or the ability to read and write, relies strictly on visual components, and ignore other forms of written communications such as Braille or electrical signals such as Morse code. But both of these examples are heavily influenced by visual written forms of spoken languages, and are sometimes direct iterations of writing systems.

1.2 Scripts and Alphabets

In his 2004 book, *The Solid Form of Language*, Robert Bringhurst says that “writing is the solid form of language, the precipitate” (p. 9). He continues, “Speech comes out of

our mouths, our hands, our eyes in something like liquid form and then evaporates at once” (p. 9, 2004). According to this definition, written materials are a detailed documentation of what we want to communicate through a series of signs and symbols.

Recognizing and understanding symbols are human beings’ abilities. According to Bringhurst, “Reading comes first. The reading of tracks and weather signs is a fundamental mammalian occupation, practised before primates started walking on their hind legs, much less using hands to write” (p. 14, 2004). He believes that we start discovering the surrounding environment through symbols, and then we apply this ability in the development of written forms of communication.

The main tool and component of literacy is the visual form of the language; we may consider writing as the visual vehicle of the spoken language. This explanation is somewhat incomplete, but it is useful as the starting point for talking about this visual form of the language. There are a variety of different approaches for visualization of spoken languages, which can be categorized. Some of the major categories of writing systems are logographic, syllabic, and alphabetic.

Peter T. Daniels, a scholar studying writing systems, asserts that writing and spoken language are fundamentally different (1996, p. 2). Daniels, following Chomsky, considers learning a language as a natural process of the human mind, but believes that literacy is a product of the human intellect that requires training. According to him, “Language is a natural product of the human mind—the properties of people that make it possible for everyone to learn any language, provided they start at a young enough age—while writing is a deliberate product of human intellect: no infant illiterate absorbs its script along with its language; writing must be studied” (1996, p. 2). This explanation puts language and literacy in two different categories. While the main purpose of writing was to communicate language, we can consider it as an artificial invention for the communication of a natural phenomenon. And as writing systems developed and evolved through time, they have played multiple roles and been dispersed into even more areas in modern societies and life within them.

The high percentage of literacy (according to the general definition of literacy) in many modern societies, along with the introduction of advanced technologies, has increased the importance of reading and writing. Moreover, knowing how to use scripts goes beyond the purpose of reading documents in modern life. We use letters to navigate rows or seats in a theater, spell an unknown word, differentiate a homonym, or clarify a word during our verbal communications. Therefore, the ability to apply the knowledge of recognizing alphabets in different places is essential for people in most societies.

Languages keep evolving and changing. Some words come to life only to become extinct, the grammatical structure and mentality of languages may change in time, and dialects keep changing through time and throughout different geographical locations. But writing systems do not follow the paths that have been taken by languages. This rigidity of writing systems became well established after the introduction of the printing press and the growth of literacy. The desire to have consistent documents across platforms and sources was the main driver for the establishment of the rules and conventions of letterforms and scripts.

Language continually develops and changes without the conscious interference of its speakers. but writing can be petrified or reformed or adapted or adopted at will. It is thus in the theoretical realm that grammatology differs most from the rest of linguistics—the theory of writing must be very different from the theory of language. It is not to be expected that patterns or principles that describe language should apply to writing, but little attention has yet been paid to that fact. (Daniels, Page 2)

But historically, alphabetic writing systems, one of which is the focus of this study, have evolved through time to meet the phonological, grammatical, and cultural needs of languages. The evolution of the Latin alphabet from the Phoenician alphabet shows the variation of shapes and numbers of letters in alphabet sets throughout history. As an example, we can look at the origin and visual representation of the letter “K”. In the Phoenician alphabet, the letter “k” was mirrored vertically (Figure 1.1). But the Greeks took the letter and turned it into the form we know today, to follow their writing direction.

According to Laurent Pflughaupt in the book *Letter by Letter*, “adopted by Greeks, who wrote from left to right starting in 550 BCE, the letter was turned to the right and given the name Kappa (κ, K)” (p. 84, 2007). In fact the letter “K” almost went extinct. Pflughaupt argues, “When the Romans invented the letter G to record a sound that was softer than C, the latter took on a harder pronunciation, causing a near disappearance of the letter K” (p. 84). He goes on to explain how Kappa was replaced by the letter C but was revived in the Middle Ages. As for writing directions, they changed over time, but there are also examples of changing direction within a single document. For example, Boustrophedon was a bi-directional type of writing in Ancient Greece. This will be explained in Section 1.4.12, Reading Direction.

Logographic Origin	Phoenician Kap	Greek Kappa
		

Figure 1.1: Evolution of the letter K

The evolution of scripts shows that certain adaptations and updates were made where required. But standardization of scripts at different time period, prevent writing systems and scripts from evolving. In comparison, spoken languages are much more flexible and less resistant to change. We may need to apply the plasticity of languages to the rigidity of writing.

1.3 The Status of the Persian Writing System

The official Persian writing system is a modified version of Arabic scripts with historical roots in the Phoenician alphabet through the Aramaic alphabet. Both the Perso-Arabic and the Roman alphabet originated from a common ancestor, the Phoenician alphabet. We can consider these two writing systems, as well as some others, including Hebrew, as branches of the same ancient writing system. The Pahlavi script, which was the

official system of the old Persia and was replaced by the Arabic script after the Muslim conquest, is more closely related to Arabic, the Kufic script, than it is to the Roman alphabet, both of which are derived from the Aramaic script. The Greeks took their alphabet directly from the Phoenicians. It is worth noting that in pre-Islamic Persia, another writing system was developed for recording the exact religious text of the time, which was called Avestan script. "The Avestan script is written in an alphabetic script created on the basis of the Middle Persian Pahlavi script" Prods Oktor Skjærvø writes, "in order to record the sacred texts precisely the way they were pronounced at the time, presumably by select priests whose pronunciation was considered to be especially reliable." (Encyclopædia Iranica, 2012)

In a way, we can consider both the ancient Greek and the Aramaic alphabets to be the grandparents of the current Roman and the Perso-Arabic alphabets. Donald Jackson explains the journey of the alphabet in *The Story of Writing*:

The alphabet developed by the Phoenicians also lies at the root of Eastern writing systems. The Aramaic peoples, of whom the Phoenicians were part, had been settled in Syria before 1000 BC, and had established themselves as prosperous and active traders throughout the area between the Mediterranean and the Euphrates valley. Their writing system was very similar to the one which the Phoenicians spread westwards, including an alphabet more or less identical to it. They were suppressed and dispersed by the Assyrian invasions after 732 BC, but they had already begun to displace the Babylonian language and cuneiform writing with their own throughout the area, and as they moved eastwards this process of linguistic colonization continued. Aramaic scripts spread throughout the Assyrian Empire, replacing cuneiform systems in the languages of Babylonia and Persia, in the lands around the Persian Gulf, and as far as Afghanistan, India and Mongolia. From these models grew the modern Arabic and Hebrew ..., as well as the non-Semitic Persian scripts, and the Brahmin script of India. (pp. 29-30, 1981)

Jackson continues on the standardization of Arabic script and relates it to the manuscript of the Koran, the religious book of Islam.

The Arabic script of Islam, which descended from the Nabatean branch of the family of Aramaic scripts, makes its first appearance around 500 AD, but early forms of it had certainly been developed at least two hundred years earlier; ... Script had been standardized by its use for the writing of the sacred Koran in the early seventh century. (Jackson, 1981)

Forogh Hashabeiky, author of *Persian Orthography, Modification or Changeover? (1850–2000)*, says

Several centuries after the Arab conquest of Iran in the seventh century, the official script of the Sasanian Empire, called Pahlavi, was replaced by Arabic script, while the spoken language continued to be Dari. Exactly when, why, and how this replacement took place is still under discussion, but the shortcomings of the Pahlavi script, the new Islamic aspirations, and the vast use of Arabic language in administration, religion, and science are among the reasons mentioned for this replacement. It has been said that the new script, still consonantal and Semitic, was welcome, as it was much easier to learn and master compared to the 14-lettered alphabet of Pahlavi script with its strong underrepresentation of sounds. (p. 17, 2005)

As stated above, Persian languages were not part of Semitic languages and cultures, nor were the Persians Semitic people. This is where Aramaic alphabets developed by these Semitic languages and cultures. According to Sadeghi and Zandi-Moghaddam in the *Persian Orthographic Dictionary* book, published by the APLL (7th edition, 2016), Persian languages are part of the Indo-European language family, while Arabic is one of the Semitic languages. As a result, these two languages have completely different structures.

The current Persian writing system was not designed for the Persian languages. Nor was the Pahlavi script, the former official writing system used in Persia. Although

Persian is not the only language that uses a writing system developed for another language, this historical inconsistency between the spoken and written forms is repeatedly addressed by scholars and linguists.

Even if the shortcomings of the Arabic alphabet were discussed by scholars from the tenth century like Ḥamzeh Eṣfahāni, as well as scholars from the eleventh century like Abu Reyḥān Biruni, the inadequacy of the Arabic alphabet in representing spoken Persian is a relatively modern concern. It has been believed [suggested] that the Semitic consonantal writing system, adopted for the synthetic structure of Semitic languages, is not suitable for an Indo-European language like New Persian with a dominating analytical structure. This belief, in its turn, is rooted in the “representational view of writing,” as Pettersson (1996:2) calls it, where the task of writing is supposed to be that of representing the spoken language, and where a one-to-one correspondence between the letters and phonemes of the language is believed to be the ideal situation. (Hashabeiky, 2005)

It's true that current Persian writing has some positive aspects that respond to the requirements of the spoken language, but the vast difficulties caused scholars like Ignace. J. Gelb to predict a shift and a transition from the Perso-Arabic writing system toward a Roman version. “Over a long period of time a language can employ several writings [systems]. Thus, for instance, the Persian language was expressed first in a variety of cuneiform writing, then in the Pahlavi and Avestan scripts, and finally in the Arabic alphabet. If and when Persia follows the example of Turkey, it is possible that we shall see Persian [Farsi] written in Latin characters” (Gelb, 1989).

Discussions about transitioning from Perso-Arabic to Roman scripts have gone on for more than 150 years in Iranian societies. Kazem Ostadi lists and expands on the motivations for transitioning the scripts in the *History of the Idea of Changing the Letters in Iran* in 2012 published by Ayene-ye-Pazhoohesh and distributed by the Center for the Great Islamic Encyclopedia as introduction of printing press to Iran, social changes and several defeats of Iranian army against Russian army during the Qajar Emperor, social

and political reforms that started during the chief minister Mirza Taghi Khan Farahani, and neighbouring with Ottoman Empire. Istanbul, considered as a geographical bridge between Eastern and Western cultures, was the major city of the Ottoman Empire, and a linking home to the Westerner modernity for Iranian.

So not only is there a history of problems with the Perso-Arabic writing system; discontent with this script system is growing in this century.

1.3.1 My Approach

Analyses of the history of Iranian and Persian scripts have been made by many researchers and scholars. Similarly, there are some references on the tendency for modernity and the idea of transitioning from Perso-Arabic scripts to the Roman scripts. Scholars like Hashabeiky (2005) and Ostadi (2012) have conducted vast and detailed research from a historical and social point of view, and we can refer to this research at any time.

The school of thought supporting the idea of changing the script system is large and full of ideas on how we can improve the writing system through this transition. At the same time there are those who think we have to stick to the existing script system.

My personal preference is for transitioning from the current writing system to a new one. But in contrast to many supporters of this change, I do not agree with the adaptation of Roman scripts for the Persian language. Although there are advantages to adopting this set of alphabets, Roman scripts impose a different set of problems on the Persian writing system. Aside from the structural problems that may be caused by adapting the Roman script, there is a possible cultural resistance against this transition influenced by cultural identity and the fear of Western hegemony of language and culture. But at the same time, there are lessons to be learned from the evolution of writing that occurred to the Greek branch of the Phoenician alphabet and that continued in the European languages and cultures, which is the better optimization of scripts for the requirements

of their languages. This is an important consideration in the case of the Persian language, because like that language, most European languages are also in the Indo-European language family, and the Persian writing system and the Roman scripts have the same ancestor, the Phoenician alphabet.

As a starting point for this research, which focuses on the role of vowels, particularly short vowels in Persian writing: If we take the evolution of writing systems as a clue, is there a potential optimization that we are missing by not having short vowels?

This question is language-dependent, to some degree. The importance, the variety, and the employment of vowels varies between languages. This paper demonstrates the relative importance of vowels in the Persian language.

Fan Fengxiang and Gabriel Altmann argue that “the grapheme-phoneme analysis of English is radically different from cases analyzed hitherto (German, Swedish, Italian, Slovak)” (2008) for three particular reasons:

- the historical origins of English,
- its many national and regional varieties and
- the borrowing of many foreign words.

It is very interesting to see that we can apply all these points to the Persian language and its writing system and analyze its grapheme-phoneme correspondence.

Historically, Persian and some European languages share common roots, but the geographical distance between these places caused a great divergence between the languages. There are many countries and societies that use Persian, along with other languages. And the current Persian language has borrowed many words from different sources, including Arabic and French.

For this research, I am using my own approach, analyzing the problems caused by the lack of short vowel alphabets in Persian writing by generating an analytic tool. First, I will expand and explain my approach, illustrating the complexity of perceiving the written

form of a word and a sentence without vowels, using English words and sentences. Then, after explaining my thought experiment, I repeat the process using Persian words and sentences, where short vowels are usually omitted. The main focus of this section is measuring the ambiguity caused by the absence of short vowels.

1.4 Terms

Before going any deeper into this research, I want to explain some ideas and terms, most of them from phonology science. These concepts are the backbone of the paper and will help me to frame the areas that I will cover later.

1.4.1 The Academy of Persian Language and Literature

The Academy of Persian Language and Literature (APLL) is the official regulatory body of and the official authority on Persian language, orthography, word formation and literature. Located in Tehran, Iran, it was founded in 1935 as the Academy of Iran, and has gone through several transitions since then. The current APLL has a close collaboration with Afghanistan and Tajikistan and has several permanent members from these countries. (APLL.ir)

1.4.2 Consonant Cluster

This term applies to a group of consonants that has no intervening vowels in between them, or to chain of consonant phonemes with no vowel phonemes. For example, the letters **rpl** in the word *purple* are a consonant cluster.

1.4.3 Diacritics

Diacritics or diacritical marks are supplementary signs added to a letter or a group of letters, usually above or below them, to indicate that the pronunciation differs from that of the original graphemes, or to indicate certain additional phones. Philip Carr (2008) describes diacritics as “a visual symbol used in transcriptions to denote some phonetic property” (p. 43).

Merriam-Webster gives a broader definition of a diacritical mark as “a mark near or through an orthographic or phonetic character or combination of characters indicating a phonetic value different from that given the unmarked or otherwise marked element” (Merriam-Webster.com, 2018).

In the Persian writing system, there are two major groups of diacritics. (Note that diacritical marks are not limited to these two groups in other script systems.) The first group is diacritical points, called *Noqte* in Persian, meaning dots or points. From one to three points are added above or below as letter to produce a new grapheme for a different phone or phoneme. Diacritics are used with various letters, up to three dots. As an example, let’s look at the following three letters, all in isolated form:

Persian Letter	Name of Letter	Number of Dots	IPA	Roman Version
ر	Re	0	/r/	R
ز	Ze	1	/z/	Z
ژ	Zhe	3	/ʒ/	ZH

Table 1.1: Diacritics make a difference to the basic shape of a letter

Table 1.1 shows the difference created by diacritical points placed above the base grapheme *Re*. By adding one dot on top of the letter *Re* /r/ (the R sound in English), the

sound changes to *Ze* /z/ (the Z sound in English). If we increase the dots to three, the sound becomes /ʒ/, like the letter S in the English word 'vision'.

The second group of diacritical marks in the Persian writing system—which is the main focus of this paper—is a group of ancillary marks or signs that are added to a letter to indicate phonemes that are not conveyed through the basic alphabet. In the Roman alphabet these diacritical marks are called accents and their function is to change the sound value of the letter to which they are added. But in Persian orthography they represent the phoneme that follows the letter they modify. This is the main difference in function between diacritical marks in the Roman-based alphabetic systems and Persian orthography. The phones that are absent in the Persian written system are three short vowels, which are called *Zebar*, *Zir* and *Pish*. *Zebar* and *Zir* are represented by small diagonal marks that are added above or below a letter. *Pish* is an apostrophe-like diacritic that is added only above a letter. These marks indicate the short vowels that are not conveyed by principal letters. This group of diacritics is explained further in Section 3.2.3.2.

There is another difference between diacritical accents in the Roman alphabet and diacritical short vowels in Persian orthography. In the Roman alphabet, an accent is added when the tone of the letter itself needs to be changed. In Persian orthography, the mark does not change the quality of the letter it modifies; it simply represents the absent sound that follows the given letter. The main points to remember are that the three short vowels are absent in Persian orthography, and these accents are usually omitted when writing in Persian.

The main difference between these two groups of diacritics is that documents without diacritical accents are usually readable, while removing diacritical points makes a document almost illegible.

The following Figure 1.2 shows a page from *Persian Writing Guideline*, a guide prepared and published by the Academy of Persian Language and Literature as an

official document on proper writing in Persian. The absence of these diacritical accents in this page is clear.



Figure 1.2: A sample page from the Persian Writing Guideline

1.4.4 Digraph

A digraph is simply a combination of two letters creating a single sound that is not represented by any single letter of the alphabet. For example, the letters s and h

together in English represent the sound /ʃ/ in IPA, like the word *ash*, which has two phonemes, /æ/ and /ʃ/, but is written with three letters. Peter Roach (2011) gives a more detailed explanation of digraphs in his book *English Phonetics and Phonology*:

It has sometimes been found necessary to combine two symbols together to represent a single sound. This can happen with alphabetic writing – the term seems mainly to be used for letter pairs in words where in Roman inscriptions the letters were regularly written (or carved) joined together (e.g. spellings such as ‘oe’ in ‘foetid’ or ‘ae’ in ‘mediaeval’), though the writing of Anglo-Saxon also involves extra symbols. It seems unlikely that anyone would call the ‘ae’ in ‘sundae’ a digraph. In the development of printed symbols some digraphs have been created, notably the combination of ‘a’ and ‘e’ in æ and ‘o’ and ‘e’ in œ; the resulting symbol when used in phonetics for vowels is supposed to signify an “intermediate” or “combined” quality.

Both homogeneous digraphs (consisting of doubled letters, like pp in *happy*) and heterogeneous digraphs (consisting of two different letters, like ch in *chair*) are permitted in English orthography. In Persian orthography, however, homogeneous digraphs are represented by a single letter, and understanding the actual meaning of the word is left up to the reader.

1.4.5 Diphthong

Diphthongs, unlike monophthongs (Section 1.4.10 Monophthong), represent a modality of transitioning from one vowel sound to another in a single syllable. Philip Carr defines a diphthong as “a vowel sound in which there is a transition from one **vowel quality** to another within a single syllable **nucleus**, as in the English word [baɪ] (*buy*)” (p. 43, 2008). “Transitions from one vowel quality to another across syllable boundaries are not diphthongs” Carr continues, “English [si:ɪŋ] (*seeing*) has a transition from [i:] to [ɪ], but this is not a diphthong, since the transition does not take place within a single nucleus”.

Diphthongs in Persian are part of the missing element of the orthography. This is explained in more detail in Section 4.4.

1.4.6 Glyph

A glyph is a visual representation or a surface form of a grapheme, as a predefined symbol and readable character. An *allograph* is a glyph that represents the same grapheme, since the possibilities of representing a grapheme are usually unlimited due to the variety of typefaces, etc. “The word glyph is used differently in different contexts. In the context of modern computer operating systems, it is often defined as a shape in a font that is used to represent a character code on screen or paper. The most common example of a glyph is a letter, but the symbols and shapes are also glyphs”

(Adobe.com)

1.4.7 Grapheme

A grapheme is the smallest component of spelling in a language and represents a phoneme. Graphemes are conceptual forms or symbols for glyphs. In the same way that phonemes are represented by phones and sounds in verbal communication, graphemes are represented by glyphs in writing. Both phonemes and graphemes are conceptual sources for phones and glyphs in two different mediums. Sometimes graphemes are represented by more than a letter or a digraph. Philip Carr describes a grapheme as “a unit in the spelling system of a language. Graphemes are usually distinguished from letters, since a given grapheme may contain more than one letter, but still act as a single unit in the spelling system. Graphemes which contain two letters are called digraphs. An example is the <th> grapheme used in English spelling, which corresponds to the phonemes /θ/ and /ð/” (2008).

1.4.8 International Phonetic Alphabet

The International Phonetic Alphabet (IPA) is an alphabetic system of representing phonetic notation developed and introduced by the International Phonetic Association in 1888. The association re-issues the IPA chart annually. According to the IPA website, “the aim of the IPA is to promote the scientific study of phonetics and the various practical applications of that science” (International Phonetic Association). IPA notations usually appear beside the original word inside a pair of brackets or between slashes. Throughout this document, I use a pair of slashes (/ /) to indicate an IPA notation; for example, the word ‘phonetic’ appears like this:

phonetic /fə'netɪk/

The 2018 edition of the IPA chart appears in Appendix A.

1.4.9 Ligature

A ligature is a typographical term for a combination of two or more letters that create a new form. The Adobe website defines a ligature as “two or more letters tied together into a single letter. In some typefaces, character combinations such as fi and fl overlap, resulting in an unsightly shape. The fi and fl ligatures were designed to improve the appearance of these characters” (Adobe.com).

In Persian orthography, letters are connected not to improve their visual appearance or typographical shape, but because the written language is strictly cursive.

1.4.10 Monophthong

A monophthong in phonology “refers to a single vowel” (Roach, 2011). It represents a pure vowel sound without any transitioning in phones. Carr describes it as “a vowel in

which the quality remains more or less constant during its production” (p. 102, 2008). A monophthong may be represented by a grapheme, like the letter *i* in the word *bit*, or by a digraph, like *oo* in the word *boot*.

1.4.11 Phoneme

A phoneme is are usually defined as the smallest unit in phonology that distinguishes words from one another other. “This is the fundamental unit of phonology” Roach writes, “which has been defined and used in many different ways. Virtually all theories of phonology hold that spoken language can be broken down into a string of sound units (phonemes), and that each language has a small, relatively fixed set of these phonemes” (2011). There are many words that have similar consonant-vowel pattern with many similar letters, with the only difference between them being a phoneme. Here is a definition and examples from Larry M. Hyman:

The phoneme was defined as a minimal unit of sound capable of distinguishing words of different meanings. Both /p/ and /b/ are phonemes in English, because they are capable of making a meaning difference, as in the words *pin* and *bin* or *cap* and *cab*. The exact interpretation of the fact that the word *pin* means something different from the word *bin* depends crucially on one's conception of what a phoneme is. (pp. 59-60, 1975)

1.4.12 Reading Direction

Scripts visually follow the directional pattern, in which they are written. Scripts are usually written horizontally (side to side) or less frequently vertically (starting at either the top or the bottom). Most modern writing systems follow the same pattern throughout the entire text. We use left-to-right (LTR) in English and other Roman-script writing systems and right-to-left (RTL) in Persian and Arabic. (Most countries in the Middle East

use this system. Turkey formerly used the Perso-Arabic script.) Some systems, like Japanese and Chinese, can be written both horizontally and vertically. In Arabic and Persian, mathematical notations and other scientific signs and symbols are written left-to-right. But this wasn't always the case. Egyptian hieroglyphs were written horizontally from both sides. Boustrophedon is a type of bi-directional writing used in ancient Greek manuscripts. As Johanna Drucker explains, "early Greek inscriptions do not have a fixed direction, they can be written right to left, or left to right, or in a regularly alternating sequence known as boustrophedon after the pattern made by a plowing ox" (p.47, 1999).

Phoenician (the ancestor of most modern alphabetic system, including the Roman alphabet and Perso-Arabic) writing in its the early stages had no standard writing direction. Andrew Robinson writes, "The direction of Phoenician writing, itself unstable prior to about 1050 BC, was stable, from right to left, probably by 800 BC.". He goes on to say, "Greeks must have borrowed the Phoenician script in the earlier phase of its development, not after it had settled down" (2007).

1.5 Scope and Limitation

Fully discussing and evaluating a writing system is a massive project, and beyond the scale of this study. In the Preface I stated my interest in the multilingual societies of the Middle East, how these societies are merged together, and the adaptation of writing systems in those places. Here I am willing to confine my interests to studying the shortcomings of the Persian writing system and their effect on the Persian language, since it is my mother tongue. To make this project feasible, I am focusing on one of several problematic characteristics of the Persian writing system, the vowel sounds. The following Figure 1.3 shows some of the problems that are more obvious for most users and critics. The focus of this paper appears highlighted in blue.

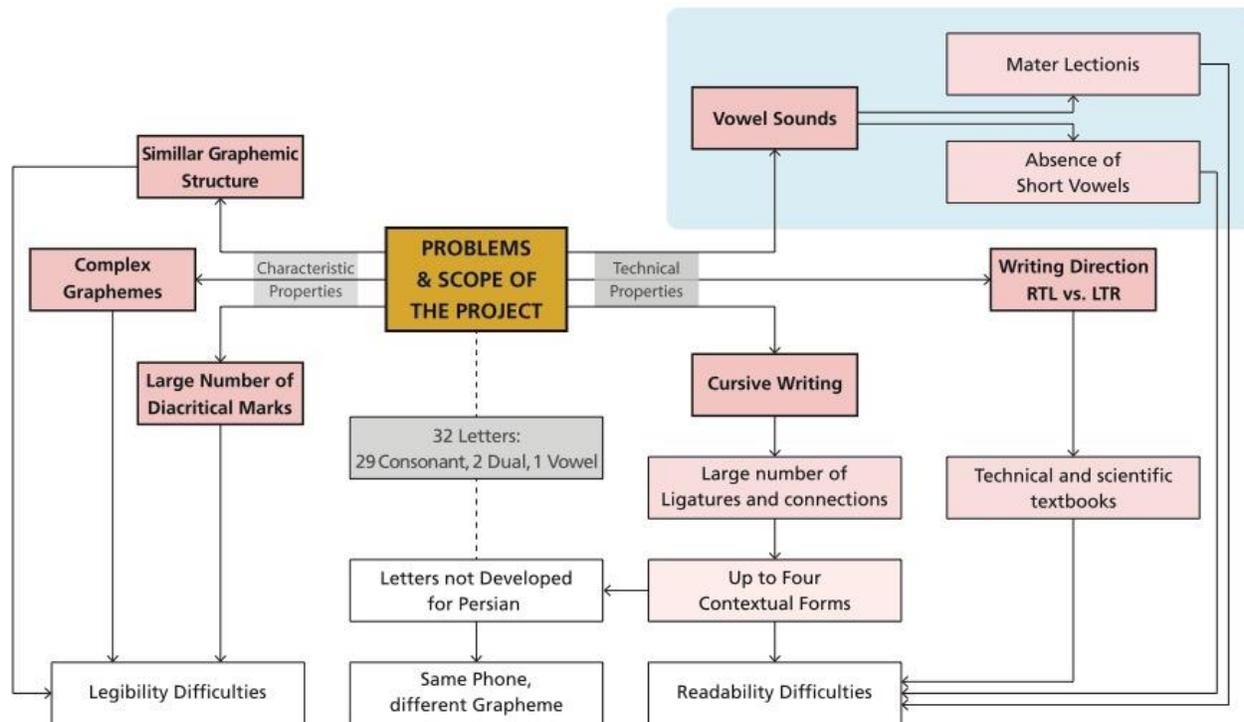


Figure 1.3: Scope of the research

1.6 Aim of This study and the Analytical Approach

The aim of this study is to demonstrate and develop an analysis of the level of ambiguity in the Persian writing system due to the lack of letterforms for short vowels and the inconsistency of vowels in general. Part of this inconsistency can be addressed by measuring the grapheme-phoneme correspondence of vowels in the Persian writing system.

To achieve the goal of this research, I am introducing an analytic approach as a tool for demonstrating this particular deficiency. With certain considerations and modifications, this approach could be used in future investigations to demonstrate different levels of ambiguity in many alphabetical writing systems.

Obviously, the lack of short vowels and the dual role of some of the long vowels are not the only problems of this script system, however, this is a good place to start.

One criticism of the analytical tool that I am introducing in this paper might be the role of context and larger documents to understanding the meaning of a given body of text.

Although I respect this point of view, but I can address this point that

- any idea or structure is made up of smaller components. Misrepresenting components of a concept or an idea allows for a high chance of miscommunication through the written form of language.
- the analytical process that is the essence of this paper focuses on the ambiguity inherent in the script and does not address the larger way in which people read
- and like other tools, this approach is designed for certain types of measurements.

1.6.1 The Argument

Although there are many factors that play a role in the average literacy rate of a society (including different notions of literacy), it is important to look at the visual components of literacy as a key element in written language. These visual components are scripts of (at least) alphabetical writing systems. One interesting aspect of the majority of alphabetical systems is that most of them have a common ancestor, the Phoenician alphabet (although language evolution has resulted in certain optimizations). Persian is one of the writing systems that evolved from the Phoenician alphabet. Its scripts were developed based on the Arabic ones, while the language itself has its roots in the Indo-European family. Prior to the Muslim conquest of Persia, people used Pahlavi scripts, which were based on the Aramaic alphabet. As a result, Persian languages borrowed from other writing systems, with different degrees of optimization, for over two millennia. Neither the Pahlavi nor the Perso-Arabic writing were perfect, or even acceptable, matches for the Persian language. The current status of the Persian writing system and language reflects its high level of inconsistency and difficulties. These problems are becoming more noticeable due to the rapidly increasing number of literate people and

the growth of digital platforms of communication. The question is whether the current writing system is able to overcome these difficulties.

1.6.2 Analytical Approach

Reports of this problem are not new. Records of this issue go back at least a few centuries, and these critiques of the writing system have become louder and more frequent in the last 150 years. There are different schools of thought on the problems of the current Persian writing system. One group of conservationist scholars prefers to keep the writing system as it is, although within this group there are those who suggest minor reforms. The other school of thought sees no future for the existing writing system and suggests either an adaptation of the Roman alphabet or the development of a new set of alphabets. Of course, there are positive and negative aspects to both of these stances.

The major criticisms of adopting the Roman scripts are

- the loss of conservation of cultural identity,
- concerns about the potential loss of the national language,
- and predominance of Western hegemony and culture

Regardless of the approach taken, the point is that any reform or alteration requires strong support and motivation. The historical overview of the current Persian language and its writing system shows inconsistencies between them. At the same time, the system's many problems —the morphology of the scripts, the phoneme-grapheme correspondence ratio, the absence of vowels, the duality in writing direction, the vast amount of diacritical points, the inconsistent cursive writing (even individual words are not always written completely in cursive) and the *Mater lectionis*— make it clear that current Perso-Arabic is not a good enough fit for the language in the modern context.

To quantify this situation, I am taking an analytic approach to find out and demonstrate the level of ambiguity created by the semi-consonantal Persian writing system. There are other areas that also need to be quantified, but those areas would require further investigation, whether as a future project for me or for anyone else with an interest in the topic.

To demonstrate the difficulties created by the lack of short vowel letters and the inconsistency of long vowel letters, I will begin first by visualizing the consequences of having no vowels in English, first at the word level and then at the sentence level. This approach gives me a quantified overview of the situation and illustrates the actual case of the Persian writing system for non-Persian readers. I will then repeat the process for Persian, with some necessary modifications. Finally, in Chapter 4, I will explain the other areas of inconsistent vowel representation in the Persian writing system.

2. Quantifying and Visualizing the Consequences of Having No Vowels at the Word and Sentence Level in English

Here in section 2, I quantify and visualize the consequences of omitting vowels in writing, using an English word and an English sentence to demonstrate the central point of this section: that a lack of vowels increases the uncertainty about the meaning of words and sentences, and this can be quantified. After demonstrating and quantifying the consequences of this ambiguity using the English writing system, the next section (Section 3) demonstrates this same point using the Persian writing system.

There are number of obvious differences between the English ant the Persian writing systems, but to demonstrate the logic of my argument, I use some English words and an English sentence to illustrate the consequences of having no vowels in a writing system. To avoid any extra complexity when visualizing the situation, I am not using words with vowel digraphs—e.g., oa, ou and ea—and am referring only to a, e, i, o, u and y (when it functions as the vowel /ɪ/, /i/ and /i:/, not the consonant /j/) while they are sitting alone in the word structures.

In this section, I examine words and sentences by removing all single vowels (a, e, i, o, u and y), and to numerate the number of possible word formation by misplacing and adding vowels in between the consonants that remain. In step one I explain the process at the word level; in step two I explain it at the sentence level.

2.1 Step 1: Word Level

At the word level, I am trying to illustrate the complexity of the written form of a word without vowels. I will use two simple words, 'bat' and 'farmer', to demonstrate and measure the possible level of ambiguity of these words. The phoneme-grapheme correspondence ratio of both of these words is 1:1, therefore they don't contain any digraphs.

2.1.1 Bat, bit or bt

I start the explanation with a very simple word, 'BAT'. Bat is a word composed of three graphemes: b, a and t, and it represents a particular sound. By removing the only vowel sound from the word, we have this structure:

b _ t

However, by arranging the letters continuously (as would be the case in Persian), we have this structure:

bt

Looking at a structure like 'bt' does not give us clear instructions for pronouncing the word, where to add vowels, or which vowels we should use. Therefore there is not an immediate meaning conveyed by 'bt'. We can potentially use any of the six vowels I mentioned above or choose not to use any vowel after each consonant. (In the example above, I am not examining any vowels at the beginning.) By adding a vowel after each letter, or skipping to add any, which gives us 72 possible formations in the 'bt' example, we have a total of 49 possible words. Table 2.1 shows the combinations.

Note:

We have to consider that there are places in words where a vowel in between consonants is not necessary. There are two major reasons for this:

- Consonant cluster is acceptable in English word structure, therefore there is not an intervening vowel in between every consonant. The letters 'mpl' in the word 'example' is a good case of consonant cluster. We may consider 'x' in the same word as a second case, as it replaces k+s and is pronounced /g/ plus /z/ in this word.
- Double letters are the other instance of English orthography that we can consider consonant clusters. The 'pp' combination in 'happy' is one example. Consonant clusters and double letters can be combined, as in the word 'battle', in which 'ttl' is a combination of both.

I use curly brackets in Table 2.1 and through this document to indicate places with no intervening vowels, or any place without vowel graphemes in between consonants, as well as the ending of the word.

Table 2.1 has seven columns, forty-nine rows and two headers. In this table, I focus on replacing the ambiguous letterforms (vowels) in the combination to demonstrate all possible interpretations of the word. Columns 1 and 4 are the locations for the letters 'b' and 't'. Column 2 is the first ambiguous letterform (vowel) placeholder for the six vowels and no intervening vowel after the letter 'b'. Column 3 is the combination of 'b' plus one of the seven possible vowels and consonant cluster in the first ambiguous placeholder, therefore there are seven cells in this column. Column 5 shows the six possible vowels with no intervening vowel coming after each of the seven cells of the third column. Column 6 shows all possible combinations for b, t and inserted vowels. Finally, Column 7 shows all meaningful combinations of Column 6.

1	2	3	4	5	6	7
b	First possible vowel insert	b+V.S	t	Second possible vowel insert	Final Combination	Meaningful Words
b	A	bA	t	A	bAtA	
				E	bAtE	BATE
				I	bAtI	
				O	bAtO	
				U	bAtU	
				Y	bAtY	
				{}	bAt	BAT
	E	bE	t	A	bEtA	BETA
				E	bEtE	
				I	bEtI	
				O	bEtO	
				U	bEtU	
				Y	bEtY	
				{}	bEt	BET
	I	bI	t	A	bItA	
				E	bItE	BITE
				I	bItI	
				O	bItO	
				U	bItU	
				Y	bItY	
				{}	bIt	BIT
	O	bO	t	A	bOtA	
				E	bOtE	
				I	bOtI	
				O	bOtO	
				U	bOtU	
				Y	bOtY	
				{}	bOt	BOT
U	bU	t	A	bUtA		
			E	bUtE	BUTE	
			I	bUtI		
			O	bUtO		
			U	bUtU		
			Y	bUtY		
			{}	bUt	BUT	
Y	bY	t	A	bYtA		
			E	bYtE	BYTE	
			I	bYtI		
			O	bYtO		
			U	bYtU		
			Y	bYtY		
			{}	bYt		
{}	b{} b	t	A	btA		
			E	btE		
			I	btI		
			O	btO		
			U	btU		
			Y	btY		
			{}	bt		

Table 2.1: Number of possible combinations of the letters 'b' and 't' using inserted vowels between the consonants

In the example 'bt', according to the instructions I gave earlier, we can form forty-nine different words, of which ten have meaning. These are:

- bate
- bat
- beta (the second letter of the Greek alphabet)
- bet
- bite
- bit
- bot
- bute (a generic name for phenylbutazone drug)
- but
- byte

If we ignored the limitations that I added to the experiment, the number of possible words increases dramatically. If we add one of the six single graphemes at the beginning, middle and end of the 'bt' combination visualized in Table 2.1, the number of possibilities is now 6×7^2 , for a total of 294. This number is even higher if we add digraph vowels. Obviously not all of the combinations are meaningful, nor do they all come to mind immediately, but the exercise illustrates the difficulty of decoding the possible combinations and the increased chance of misrepresenting a concept throughout the written form of the language. Because alphabet scripts represent spoken language, it is worth noting that other words appear here that phonetically represent English words such as bote ("boat") and bete ("beet"), though, as digraphs in English orthography, these words are not part of this list.

2.1.2 Farmer frmr

The next word that I am using as an example is 'farmer': a word comprised of four consonants and two vowels (letter r in this word effects the pronunciation of vowels) with the following pattern, based on consonants (C), vowels (V), and no intervening vowels ({} in between consonants or at the end of the word:

CVC{}CVC{}

After removing the vowels from the word, we are left with the following letters:

f_rm_r

By deriving just the consonants from the word, we have the following combination:

frmr

I am repeating the same exercise I did in part 2.1.1 with the word 'bat'. Here, there are four potential locations in which to insert one of the six vowels (a, e, i, o, u and y), or where we can omit vowels in between the consonants (consonant cluster) or at the end of the word.

f □ r □ m □ r □

There are four possible places for each of the six vowels or consonant cluster (seven options). The number of possible word combinations is 7^4 , a total of 2,401. If we start the word with a vowel, then the number of combinations increases to 6×7^4 , or 14,406. Again, not all of these possible words have meaning, nor are there immediate options for the 'frmr' consonant skeleton to form a word, but the number is significant.

Here are three possible immediate word combinations with similar CVCCVC patterns:

- farmer
- former
- firmer

And a less immediate word with a CCVCVC pattern:

- framer

Perceiving and understanding a visually represented word (whether as a standalone element of language or as a piece of a sentence) is a complex process consisting of properties like memory and the literacy level of a reader, and we cannot measure the entire process with simple math. But the main concept of these numbers is to visualize the increased level of ambiguity of having no vowels in reading and writing in an alphabetic language.

There are two other outcomes in this section to analyzing these two examples at the word level:

- The number of new word possibilities increases based on the number of consonants in a word, or any consonant combination.
- The number of meaningful words derived from the initial consonant combinations are not directly related to the number of consonants. While fewer consonants may yield more possible meaningful words, more consonants may lead to less meaningful outcomes.

2.2 Word patterns

By breaking down words into smaller pieces, we can address two main categories of phonemes: consonants and vowels. The majority of English words are composed of a

combination of both of these elements. With these two types of phonemes, we can form a variety of patterns, and make a variety of words by replacing each place holder (Consonant or Vowel) with an actual phoneme. We use graphemes or letters to demonstrate these phonemes. We have to note that the lack of phoneme-grapheme correspondence increases the complexity of word composition, which I will explain in more detail after the provided examples.

Here I will explain word patterns based on consonant sounds priority, then I will form patterns by adding vowels. The purpose of this section is to examine and to understand the role of written vowels in words and sentences. All patterns are based on consonants with a 1:1 ratio of phoneme-grapheme correspondence to avoid increasing the level of complexity.

2.2.1 Words with only one consonant

The simplest and shortest form of a word, based on the consonant sound priority, is a word that is composed of only one consonant. In Table 2.2, I explain the number of possible combinations and give an example for each pattern. The three top rows of the table are the compositions based on single graphemes, both consonants and vowels. In the rest of the table, single consonant graphemes are combined with either single grapheme vowels or digraph vowels.

Vowel	Vowel	Consonant	Vowel	Vowel	Example
	V	C			Up, On
		C	V		No, Go
	V	C	V		Era, Ago
V	V	C			Out, Aim
		C	V	V	Due, Low
V	V	C	V	V	Audio
V	V	C	V		Easy, Auto
	V	C	V	V	Area, Idea

Table 2.2: Word patterns based on one consonant and up to two vowels

Based on Table 2.2, we can form these eight patterns with a single consonant:

- VC
- CV
- VCV
- VVC
- CVV
- VVCVV
- VVCV
- VCVV

Therefore, any consonant fits into one of these eight patterns.

We can study the word formation and patterns just by checking the phonemes, and leave the phoneme-grapheme correspondence for a later step. Here is the top part of Table 2.2 as reference:

Vowel	Consonant	Vowel
V	C	
	C	V
V	C	V

Table 2.3: Possible word patterns for a single consonant phoneme and vowel graphemes

According to Table 2.3, there are three possible patterns for forming a word, if we analyze word formation just by looking at the possible phonemes.

2.2.1.1 Vowel phonemes

There are 26 letters in the formal English writing system to represent any word and phoneme in the language, but the number of phonemes is higher than the number of available letters. Therefore, there are a number of digraphs in English writing for both vowels and consonants. The number of these phonemes is variable based on different dialects from various societies and countries. But Carnegie Mellon University developed an open-source pronunciation dictionary for North American English called *The CMU Pronunciation Dictionary* or *CMUdict*. CMUdict determined that there are 39 phonemes, including 15 vowel phonemes. Table 2.4 show the CMUdict phoneme table (*CMU Pronouncing Dictionary*, 2014) with highlighted vowel phonemes:

Phoneme	Phones	Example	CMUdict Translation
AA	vowel	odd	AA D
AE	vowel	at	AE T
AH	vowel	hut	HH AH T
AO	vowel	ought	AO T
AW	vowel	cow	K AW
AY	vowel	hide	HH AY D
B	stop	be	B IY
CH	affricate	cheese	CH IY Z
D	stop	dee	D IY
DH	fricative	thee	DH IY
EH	vowel	Ed	EH D
ER	vowel	hurt	HH ER T
EY	vowel	ate	EY T
F	fricative	fee	F IY
G	stop	green	G R IY N
HH	aspirate	he	HH IY
IH	vowel	it	IH T
IY	vowel	eat	IY T
JH	affricate	gee	JH IY
K	stop	key	K IY
L	liquid	lee	L IY
M	nasal	me	M IY
N	nasal	knee	N IY
NG	nasal	ping	P IH NG
OW	vowel	oat	OW T
OY	vowel	toy	T OY
P	stop	pee	P IY
R	liquid	read	R IY D
S	fricative	sea	S IY
SH	fricative	she	SH IY
T	stop	tea	T IY
TH	fricative	theta	TH EY T AH
UH	vowel	hood	HH UH D
UW	vowel	two	T UW
V	fricative	vee	V IY
W	semivowel	we	W IY
Y	semivowel	yield	Y IY L D
Z	fricative	zee	Z IY
ZH	fricative	seizure	S IY ZH ER

Table 2.4: The CMUdict phoneme table

We can update Table 2.3 with the possible number of vowel phonemes based on the CMUdict. Following is the updated table:

Vowel	Consonant	Vowel	Possible number of combinations
V	C		15
	C	V	15
V	C	V	$15^2=225$
Total:			255

Table 2.5: Number of possible word combinations with the CMUdict vowel phonemes

Table 2.5 shows that there are 255 possible combinations for a single consonant phoneme.

2.2.1.2 Digraph and diphthong mapping

While Table 2.5 is all about phonemes and pronunciations of a possible word, the next step is writing words with only twenty-six letters, while the number of phonemes is higher than twenty-six. Due to a different number of available phonemes and available graphemes in North American English, using multiple letters to represent unavailable graphemes is inevitable. These letter combinations or digraphs happen in both consonants and vowels, such as in combinations like ch to represent /tʃ/, and sh to represent /ʃ/ in consonant phonemes, and oy to represent /ɔɪ/. Not all vowel digraphs represent different phonemes or diphthongs, nor are these combinations mapped entirely differently, but there are some combinations that represent the same sound; for example, ‘a’ in ‘fake’, and ‘ay’ in ‘pay’:

fake /feɪk/

pay /peɪ/

In the first word, 'pay', we have a diphthong after 'p', represented by a letter pair. In the second word, 'fake', we have the same diphthong, but it is now represented by a monographeme or letter. There are examples of the opposite, where we have a monophthong that is sometimes represented by a single grapheme and other times with a digraph. For example, the phoneme /ʊ/ is represented by the letter 'u' in 'put' and by a double 'o' in 'foot'. A third situation is that of a single grapheme, which is not paired with any other vowel grapheme but plays two different roles in two different situations. The letter 'u' in 'pure' functions as a diphthong, while in 'pull' it is a monophthong.

If we pick, for example, the single letter **b**, one of the first obvious words that comes to mind is "be", with the CV pattern. But we can easily spot these words with the CVV pattern:

- buy
- bye
- boy
- bay
- bio
- bow
- bee

Considering the usage of two letters to represent a single vowel sound or diphthong expands the number of possible combinations shown in Table 2.5 from CV, VC and VCV to the following table, which is a different version of Table 2.2.

Vowel	Vowel	Consonant	Vowel	Vowel
v	v	C		
		C	v	v
v	v	C	v	v
v	v	C	v	
	v	C	v	v

Table 2.6: A single consonant letter in combinations of digraphs and single vowel letters

There are some letter pairs that represent either a monophthong or a diphthong with the combination of a single vowel with ‘w’ and ‘y’. CMUdict considers these two letters as a semivowel (CMU Pronouncing Dictionary, 2014), which is reflected in Table 2.4. And these two letters expand the number of word possibilities. Obviously, we cannot form an infinite number of meaningful words, and there are rules about and exceptions to pairing and combining letters to write and form a word.

The following table 2.7 shows parts of possible letter pairings used to create vowel digraphs:

	a	e	i	o	u	y	w
a		ae	ai			ay	
e	ea	ee	ei		eu	ey	ew
i		ie					
o	oa	oe	oi	oo	ou	oy	ow
u		ue	ui				

Table 2.7: Vowel digraphs of the English writing system

In Table 2.7, there are 19 pairs of letters to represent monophthong and diphthong vowels. With these 19 vowel pairs and the single vowels from Table 2.1, we can update Table 2.5 to measure the number of possible word combinations that include both monophthongs and diphthongs. Table 2.8 is an updated version of possible word combinations using a single consonant.

Vowel	Consonant	Vowel	Possible number of combinations
V	C		25
	C	V	25
V	C	V	$25^2=625$
Total:			675

Table 2.8: Number of possible words based on any kind of vowel sound using a single consonant

Therefore, by having a single consonant letter, we can form a word based on one of the previous patterns. These vowel placeholders give three options: single vowel graphemes, vowel digraphs and diphthongs. For forming a word from any single consonant, we have two steps: finding the right pattern and using the right vowel phoneme to add to the patterns. The final number for making a single consonant word is 675. By adding other letter pairs from loan words, this number becomes even higher.

2.2.2 Words with two consonants

The second obvious choice for a word based on the consonant priority is a word with two consonants. As in section 2.2.1, here I talk about the possible combination of consonants with the 1:1 phoneme-grapheme correspondence, to avoid additional consonant complexity. Table 2.9 shows different word patterns based on the variety of consonant and vowel combinations. In this table I considered both monophthongs and diphthongs as vowel sounds.

Vowel	Consonant	Vowel	Consonant	Vowel
V	C		C	
V	C	V	C	
V	C	V	C	V
V	C		C	V
	C	V	C	V
	C	V	C	

Table 2.9: Possible word patterns using two consonant phonemes

Based on Table 2.9, there are six possible formations for a word with two consonant phonemes. Referring to Table 2.7 plus the six single vowels at the start of this chapter, we can assume 25 possibilities for each vowel cell in Table 2.9. The result is then added to Table 2.10.

Vowel	Consonant	Vowel	Consonant	Vowel	Possible number of combinations
V	C		C		25
V	C	V	C		$25^2=625$
V	C	V	C	V	$25^3=15,625$
V	C		C	V	$25^2=625$
	C	V	C	V	$25^2=625$
	C	V	C		25
Total:					17,550

Table 2.10: Number of possible word combinations based on two consonant sounds

The total number of possible word combinations using two consonants is 17,550. Obviously, there are not 17,550 meaningful words for any pair of consonants; at the

same time, there are lots of additional rules and exceptions for forming a word in English. We can add the number of extra words that do not fit into this chart, including loan words.

The following is a list of some possible words made using the letters 'm' and 'n' sequentially, following one of the patterns in Table 2.9:

- man
- men
- menu
- mono
- main
- mean
- moon
- moan
- mine
- many
- money
- mania

Increasing the number of consonant sounds raises the possibilities of forming a word mathematically, but these numbers are not the only considerations in forming, writing and understanding a word. words come into language from a variety of historical, cultural and even geographical sources. There are phonemes, both consonants and vowels, that have no equivalent when a word introduces to a new language, and there are other phonemes that are pronounced differently in different branches and dialects of the same language.

2.3 Sentence Level

We use language to communicate different concepts, from the simplest of meanings to very abstract implications. Therefore communication goes beyond just words. We form sentences, paragraphs and longer word chains. And based on the previous parts of this section, if we consider the absence of vowel sounds in various words, this can create great uncertainty, ambiguity and difficulty when reading a sentence.

To explain this situation, I use the following sentence to illustrate the mathematical form of the ambiguity:

'Tom lost his hat in the market.'

I chose these words without vowel digraphs to follow a similar choice that I made in the word level of this section. I have removed all of the vowels from the sentence to produce 't_m l_st h_s h_t _n th_ m_rk_t'.

By writing the consonant skeleton of the sentence we end up with the following:

tm lst hs ht n th mrkt.

This sentence contains seven words, ranging from one to four consonants. In Sections 2.2.1 and 2.2.2, you can find the possible patterns of single and double consonant words. Below are two tables demonstrating the possible patterns for three- and four-consonant words. To avoid extra complexity in these tables, I assume a maximum of two continued consonants without an intervening vowel.

Vowel	Consonant	Vowel	Consonant	Vowel	Consonant	Vowel	Possible number of combinations
V	C		C	V	C	V	$25^3=15,625$
V	C		C	V	C		$25^2=625$
	C		C	V	C	V	$25^2=625$
	C		C	V	C		25
V	C	V	C		C	V	$25^3=15,625$
V	C	V	C		C		$25^2=625$
	C	V	C		C	V	$25^2=625$
	C	V	C		C		25
V	C	V	C	V	C	V	$25^4=390,625$
V	C	V	C	V	C		$25^3=15,625$
	C	V	C	V	C	V	$25^3=15,625$
	C	V	C	V	C		$25^2=625$
Total:							456,300

Table 2.11: Number of possible word combinations based on three consonant sounds

Vowel	Consonant	Vowel	Consonant	Vowel	Consonant	Vowel	Consonant	Vowel	Possible number of combinations
V	C	V	C	V	C	V	C	V	$25^5=9,765,625$
V	C	V	C	V	C	V	C		$25^4=390,625$
	C	V	C	V	C	V	C	V	$25^4=390,625$
	C	V	C	V	C	V	C		$25^3=15,625$
V	C		C	V	C	V	C	V	$25^4=390,625$
V	C		C	V	C	V	C		$25^3=15,625$
	C		C	V	C	V	C	V	$25^3=15,625$
	C		C	V	C	V	C		$25^2=625$
V	C	V	C		C	V	C	V	$25^4=390,625$
V	C	V	C		C	V	C		$25^3=15,625$
	C	V	C		C	V	C	V	$25^3=15,625$
	C	V	C		C	V	C		$25^2=625$
V	C	V	C	V	C		C	V	$25^4=390,625$
V	C	V	C	V	C		C		$25^3=15,625$
	C	V	C	V	C		C	V	$25^3=15,625$
	C	V	C	V	C		C		$25^2=625$
V	C		C	V	C		C	V	$25^3=15,625$
V	C		C	V	C		C		$25^2=625$
	C		C	V	C		C	V	$25^2=625$
	C		C	V	C		C		25
Total:									11,846,900

Table 2.12: Number of possible word combinations based on four consonant sounds

Tables 2.11 and 2.12 demonstrate the number of possible word combinations based on three and four consonants. In Table 2.13, I will focus on the number of possibilities in the sentence.

Consonant combinations	Number of patterns	Number of possible words
tm	6	17,550
lst	12	456,300
hs	6	17,550
ht	6	17,550
n	3	675
th	6	17,550
mrkt	20	11,846,900
Total:		12,374,075

Table 2.13: Number of word combinations and possibilities for the sentence based on individual words

In Table 2.13, each word is placed in the first column. The second column expresses the number of possible consonant-vowel combinations. Finally, the third column shows the number of possible word formations. At the end, we have an enormous number of possibilities for the sentence. The number of possible meaningful words for each consonant group is much smaller than the figure that appears in the third column. At the same time, the consonant pairs of similar length form a different number of meaningful words. The next reduction is to make a sentence, or meaningful sentences, from the sentence level example.

In the same way that not all possible letters can make up a meaningful word, meaningful words, even when grouped together, do not always form a sentence either grammatically or rationally.

The grammatical function of a word is another important aspect of delivering meanings in sentences. For example, in the given sentence, the 'tm' combination could represent three different names and persons. These persons, and the importance of losing

information due to uncertainty recognizing the right individual, is more important in this example compared to 'n', which represents either 'on' or 'in'.

2.4 Implications of having no vowels in English orthography

We know that English orthography has a variety of single vowels and vowel digraphs to represent monophthongs and diphthongs. And this section is all about visualizing and understanding the Persian orthography case, in which vowels are not fully represented in the written form of the language. A language's orthography is a series of rules and irregularities meant to code the meanings and concepts of authors and allow readers to decode written material. Writing and reading are complex cognitive processes of coding and decoding meanings through series of symbols. All this means that representing the orthography of a language quantitatively requires additional layers and considerations.

In the next section, I will demonstrate and explain the actual situation of Persian orthography by addressing the partial appearance of vowel sounds; the dual roles of some vowel graphemes (as vowel and consonant); and the way Persian writers omit parts of vowels while using the Perso-Arabic alphabet.

3. Quantifying and Visualizing the Consequences of Absence of Short Vowel Letters in the Persian Writing System

In Section 2, I explained and visualized a simplified version of the consequences of omitting vowel sounds in English writing to demonstrate my argument about the ambiguity the Persian writing system faces due to phoneme-grapheme discordance.

First, I introduce the Persian writing system very briefly, then compare it to English to clarify some similarities and differences. And finally, I demonstrate how the lack or inconsistency of vowels in the Persian writing system, causes difficulties for people who use it.

3.1 Similarities and differences between the Persian and English writing systems

There are similarities in the architecture of the Persian and Roman alphabets. These come from a common ancestor: the Phoenician alphabet. The main commonality between these writing systems is that they are alphabetic systems. Both of these systems are written horizontally on a main baseline, with scripts drawn above and below it up to a certain distance. A baseline is called (*Khat-e-Korseh*) or (*Khat-e-Zameene*) in Persian, and Persian is written from right-to-left. Figure 3.1 shows both systems, the

Persian scripts on top and the English version of the Roman alphabet on the bottom, with the baseline highlighted in blue. The Persian sample is part of a short story by Bozorg Alavi from *Varaq Pāreh'hāye Zendān* [*Scrap Papers from Prison*], originally published in 1941, which I use throughout this paper as a longer sample text. The English sample is part of Bertolt Brecht's "A Short Organum for the Theatre", originally published in 1949, from *Brecht on Theatre: The Development of an Aesthetic*, translated and edited by John Willett and published in 1978.

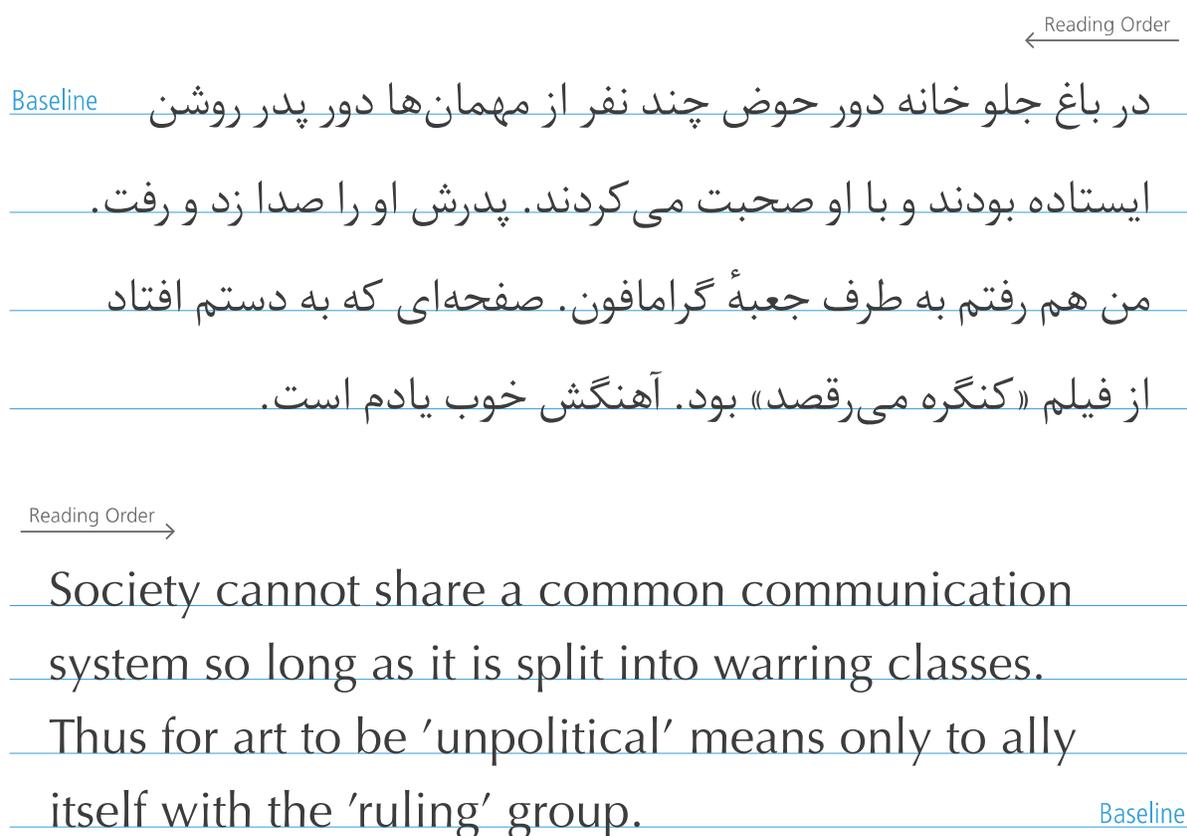


Figure 3.1: Comparing two writing systems, the Persian and the English

At the same time, the differences between these systems are more obvious than the similarities. These differences are the result of the distinctive paths the writing systems evolved over time in order to adapt to their respective languages and cultures.

Three of the main differences between the systems are the reading direction, the treatment of individual versus cursive letters, and the morphology of the letters. English

is written from left-to-right, while Persian is written the opposite way, from right-to-left (however, in Persian, mathematical notations are written left to right, although with a different set of numerical digits compared to the universal set of numbers). In written words, most of the Persian letters are attached to one another without a break, whereas Roman letters are written separately (unless cursive script is being used). And finally, the overall shapes of these two alphabet sets are different. Figure 3.1 shows the reading direction in both systems.

3.2 The Persian alphabet and writing system

The Persian writing system is a set of scripts, written cursively from right-to-left, derived from the Phoenician alphabet through the Arabic script after the Arab conquest of Persia. The Persian alphabet is called *Alefbā* /ælefbɑː/, which has a similar pronunciation to the English word ‘alphabet’.

Peter T. Daniels introduced the word *Abjad* for the Arabic writing system, in his book *The World’s Writing Systems* and explained it as a consonantal alphabet.

In a consonantary, here called an abjad as a parallel to “alphabet” (the word is formed from the first letters of the most widespread example, the Arabic script, in their historic order), the characters denote consonants (only). In an alphabet, the characters denote consonants and vowels. (Daniels, 1995)

Based on this explanation, the Persian Alefbā is a modified abjad, in which long vowels are normally written while short vowels are omitted. But in the same book, on page 747, Daniels categorized Persian writing as simply Abjad. Persian does not fit the above definition of Abjad as a consonantary script, since long vowels are represented in Persian writing.

The Persian alphabet is a set of 32 letters (some sources consider it to contain 33 letters because they include an Arabic grapheme that is sometimes incorporated into Persian writing). Each letter has two to four contextual forms: **initial**, **medial**, **final** and **isolated**, which I will explain in more detail in section 3.2.2. Table 3.1 shows the Persian alphabet chart, with all possible contextual forms of each letter, the equivalent Roman letters, IPAs (International Phonetic Alphabets), Unicodes, and DIN (Deutsches Institut für Normung) 31635 standard. The small arrows in the header of the table under the Persian name of the letters and the contextual forms indicate the reading direction of the columns.

There are few other sources available to represent phonetics and Romanize the Persian scripts, like the Library of Congress standard for Romanization the Persian writing system. But adding more sources makes the table more complex. For the Roman equivalent of the Persian letters, I am using the common scripts that most Persian-speaking people use for text messaging, unofficial emails, social networks and other digital platforms where the Persian scripts is not possible or not the first choice. This transliteration is not aligned with most of the standards. I extracted the Persian IPA from the 2015 revision of the International Phonetic Alphabet for the table and the entire paper. For the Unicode column I used version 11 of the Unicode Standard, range: 0600-06FF for the Arabic language. And finally, I used the 2011 edition of the Deutsches Institut für Normung (German Institute for Standardization, or DIN), standard 31635, for information and documentation, Romanization of the Arabic Alphabet for Arabic, Ottoman-Turkish, Persian, Kurdish, Urdu and Pashto, for the DIN column.

No.	Name		Contextual Forms				Roman	IPA	Unicode	DIN 31635
	Pe.	En.	Iso.	Ini.	Med.	Fin.				
	←		←							
1	الف	Alef	ا	ا	ا	ا	A	/ɑ/ /ɑː/ /æ/	U+0627	ā
2	ب	Be	ب	ب	ب	ب	B	/b/	U+0628	b
3	پ	Pe	پ	پ	پ	پ	P	/p/	U+067E	p
4	ت	Te	ت	ت	ت	ت	T	/t/	U+062A	t
5	ث	Se	ث	ث	ث	ث	S	/s/	U+062B	ṣ
6	جيم	Jim	ج	ج	ج	ج	J	/dʒ/	U+062C	j
7	چ	Che	چ	چ	چ	چ	CH	/tʃ/	U+0686	č
8	ح	He	ح	ح	ح	ح	H	/h/	U+062D	ḥ
9	خ	Khe	خ	خ	خ	خ	KH	/x/	U+062E	ḫ
10	دال	Dāl	د	د	د	د	D	/d/	U+062F	d
11	ذال	Zāl	ذ	ذ	ذ	ذ	Z	/z/	U+0630	ẓ
12	ر	Re	ر	ر	ر	ر	R	/r/	U+0631	r
13	ز	Ze	ز	ز	ز	ز	Z	/z/	U+0632	z
14	ژ	Že	ژ	ژ	ژ	ژ	Ž	/ʒ/	U+0698	ž
15	سين	Sin	س	س	س	س	S	/s/	U+0633	s
16	شين	Shin	ش	ش	ش	ش	SH	/ʃ/	U+0634	š
17	صاد	Sād	ص	ص	ص	ص	S	/s/	U+0635	ş
18	ضاد	Zād	ض	ض	ض	ض	Z	/z/	U+0636	ẓ

No.	Name		Contextual Forms				Roman	IPA	Unicode	DIN 31635
	Pe.	En.	Iso.	Ini.	Med.	Fin.				
	←		←							
19	طا	Tä	ط	ط	ط	ط	T	/t/	U0637+	ṭ
20	ظا	Zä	ظ	ظ	ظ	ظ	Z	/z/	U0638+	ẓ
21	عین	Ein	ع	ء	ع	ع	'	/ʔ/	U0639+	‘
22	غین	Ghein	غ	غ	غ	غ	GH	/ɣ/ /ɣ̣/	U063+A	ğ
23	ف	Fe	ف	ف	ف	ف	F	/f/	U0641+	f
24	قاف	Ghäf	ق	ق	ق	ق	GH	/ɟ/	U0642+	q
25	کاف	Käf	ک	ک	ک	ک	K	/k/	U06+A9	k
26	گاف	Gäf	گ	گ	گ	گ	G	/g/	U06+AF	g
27	لام	Läm	ل	ل	ل	ل	L	/l/	U0644+	l
28	میم	Mim	م	م	م	م	M	/m/	U0645+	m
29	نون	Nun	ن	ن	ن	ن	N	/n/	U0646+	n
30	واو	Väv	و	و	و	و	V	/v/ /u:/ /o/ /ow/	U0648+	v / ū / ow
31	ه	He	ه	ه	ه	ه	H	/h/	U0647+	h
32	ی	Ye	ی	ی	ی	ی	Y	/j/ /i/ /ɔ:/	U06+CC	y / ī / á
Pe.: Persian En.: English Iso.: Isolated Ini.: Initial Med.: Medial Fin.: Final										

Table 3.1: The Persian alphabet chart, contextual forms and pronunciations

Because Persian writing is cursive, depending on the position of the letter in a word and the prior and following letters, one of these contextual forms should be used. Therefore, a letter's shape can vary in a word. Cursive writing is explained in more detail in section 3.2.2.

3.2.1 Diacritical marks

Diacritical marks are an essential part of the Persian alphabet (see 1.4.3 *Diacritics too*). Lots of letters are distinguished by diacritical marks such as points, therefore there are similarities between letters despite the phoneme they represent. Diacritical points and diacritical signs can be located above or below a letter, regardless of whether they are permanent or temporary parts of the letter. For example, the letters *Be* (the sound B in English orthography) and *Te* (the sound T in English orthography) have a very similar structure but represent completely different phonemes, while the only difference between these letters is the number and location of the diacritical points. The following figure, Figure 3.2, shows these two letters in isolated form with highlighted diacritical points.

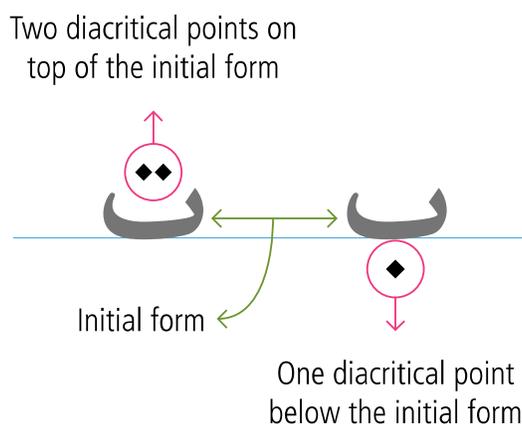


Figure 3.2: Letters Be and Te

The letters ر, ز, ژ are another example where a foundational form can be used to create three distinct letters by adding diacritical points. These three letters are formed based on the letter ر /r/. The first letter “ر”, called “Re” in Persian, is the initial form. The second one “ز” (Ze), sounds like ‘Z’ in English and /z/ in the International Phonetic Alphabet

(IPA; explained in section 1.4.8). It is distinguished by one diacritic on top of the letterform “Re”. Finally, the letter “ژ” (*Zhe*), / ʒ / in IPA, which sounds like the “s” in ‘measure’ is derived from the letter “Re” with three additional diacritical points.

3.2.2 Cursive writing

Due to the cursive character of written Persian, the form of each letter depends on the letter that comes before it and the letter that falls after it. Therefore there are up to four contextual forms for each letter: initial, medial, final and isolated. Here is the explanation and the role of each of these forms:

- Initial: the letter joins the subsequent letter from the left. This may occur at the beginning or in the middle of a word.
- Medial: the letter joins (in between) the preceding and following letters from both sides. This form of letter only occurs in the middle of a word.
- Final: the letter only joins the preceding letter from the right side. This form of letter may occur in the middle of a word or at the end of it.
- Isolated: the letter does not join any preceding or following letters. This form of letter may come at the beginning, in the middle, or at the end of a word.

The following Figure 3.3 is an excerpt from the same source as figure 3.1, by Bozorg Alavi, to illustrate the regular Persian writing:

در باغ جلو خانه دور حوض چند نفر از مهمان‌ها دور پدر روشن
 ایستاده بودند و با او صحبت می‌کردند. پدرش او را صدا زد و رفت.
 من هم رفتم به طرف جعبه گرامافون. صفحه‌ای که به دستم افتاد از
 فیلم «کنگره می‌رقصد» بود. آهنگش خوب یادم است. ولی شعرها از
 یادم رفته است. ترجمه آن به فارسی چیزی شبیه به این بوده است:
 «یک مرتبه فقط پیش آمد می‌کند، دفعه دوم ندارد. آنچه را که زندگی
 یک مرتبه به ما پیشکش کرده است، دیگر نمی‌دهد، دیگر نمی‌دهد.»

Figure 3.3: An example of Persian writing

Table 3.1 shows all letters and contextual forms. I am using letter *Be* with isolated form 'ب' to demonstrate all four contextual forms in words. Letter *Be* sounds like the letter B in English orthography and /b/ in IPA.

In the following Table 3.2, the letter *Be* is highlighted in red.

Contextual Form	Persian Word	Meaning	IPA	Roman
Initial	باد	Wind	/bɒ:d/	Bād
Medial	سبز	Green	/sæbz/	Sabz
Final	سیب	Apple	/si:b/	Seab
Isolated	آب	Water	/ɒ:b/	Āb

Table 3.2: Letter Be in different contextual forms

Additionally, there are seven letters that do not connect to the letters that come after them, regardless of word positioning. This is why some words exist as a unit, as though with a single ligature (ligature in typography is a situation when two or more letters are combined to create a single glyph, like f+i, which makes fi), whereas other words consist of isolated forms. In figure 3.4, words are marked with blue lines, and connections and disconnections are clear in the middle of each word. At the same time, words without any disconnection are marked with green lines.

آهنگش خوب یادم است. ولی شعرها از یادم رفته است. ترجمه آن
 به فارسی چیزی شبیه به این بوده است: «یک مرتبه فقط پیش آمد
 می کند، دفعه دوم ندارد. آنچه را که زندگی یک مرتبه به ما پیشکش
 کرده است، دیگر نمی دهد، دیگر نمی دهد.»

Figure 3.4: Highlighted words in two groups

Figure 3.5 gives the seven letters that can only connect to preceding letters.

ا	د	ذ	ر	ز	ژ	و	at the beginning, middle and end of words, without any connection from either side
ا	د	ذ	ر	ز	ژ	و	in the middle and end of words, with only connection from right side

Figure 3.5: The seven Persian letters that only join preceding letters

In addition, not all letters connect to their preceding and following letters. This is another variable in choosing the right form of a letter in a word. Because of this characteristic of the Persian alphabet, all words follow one of these three possible formats:

- words are a block unit and letters are connected
- words are formed with both connected and individual letters
- words contain only individual letters

Table 3.3 shows an example of each of these three situations:

Row	Persian Word	Meaning	IPA	Roman
1	سبز	Green	/sæbz/	Sabz
2	قرمز	Red	/ger'mez/	Ghermez
3	زرد	Yellow	/zærd/	Zard

Table 3.3: Three Persian words composed differently in terms of cursive writing

The word in Table 3.3, row 1, is composed of connected contextual forms. The second word, in row 2, is composed of a mixture of contextual forms, and as a result there are connections and disconnections in one word. And the third word, in row 3, is a word composed completely of isolated letterforms.

Each of these examples represent the name of a colour in Persian. Examples one and three are written with three consonants in Persian, but both have a short vowel after the first consonant. The phonetic pattern of words in row 1 and 3 is:

CVCC

but in writing we have the following pattern with an omitted short-vowel grapheme:

CCC

The word in row 2 is written with four consonant graphemes, and its phonetic pattern is CVCCVC, while its written pattern represents CCCC with two omitted short-vowel graphemes.

3.2.3 Vowel sounds

Vowels, the main focus of this research, add another layer of complexity to the Persian writing system. There are two major types of vowel sounds in Persian: short vowels and long vowels. Short vowels are omitted, unless certain diacritics are added to them to avoid ambiguity in complex words, or with specialized vocabulary such as what is used in law or medicine. Unlike short vowels, which are generally omitted, long vowels are commonly used in Persian writing.

3.2.3.1 Long vowels

There are three major long vowels in Persian: 'آ, و, ی'.

The first long vowel, which is the first letter of the Persian alphabet, is **Alef**. It has two isolated forms, "ا" and "آ", depending on the location of the letter in a word. In the word "باد" /bɑ:d/ (baad in Roman transliterated form), meaning "wind", the letter Alef stands in between two consonants. The pattern for this word is CVC.

The second long vowel is **Vāv** (/vɑ:v/), with an isolated form of "و". An example using this letter is "روز" /ruz/ (rouz in Roman transliterated form), meaning 'day'. As with the word "باد" (baad) in the previous example, it has the pattern CVC.

And finally, we have the long vowel **Ye**, with an isolated form of "ی". An example with a similar CVC pattern for this letter is "میز" /mi:z/ (meaz in Roman transliterated form) meaning "desk". In this example, the letter Ye appears in the medial form.

3.2.3.2 Short vowels

There are three short vowels in Persian, '/æ/, /e/ and /o/'. Short vowels are not part of the alphabet set, and as stated above, these phonemes are omitted from the Persian writing system. (The reason for this is that this group of vowels letters are the continuation of the vowel-less Phoenician writing system, which was introduced to

Persian writing through the Arabic language). Persian words are unaccented. The Academy of Persian Language and Literature allowed the elimination of accents. (There are some cases where accents are still used: to simplify complicated or unknown words; to distinguish similar words; or to clarify an accent that is different from the official Persian).

The first short vowel phoneme /æ/, called **Zebar**, is a small diagonal accent that appears on top of a letter (َ). It sounds like the letter ‘a’ in the word “bat”. In the word “برگ” /bærg/ (barg in Roman transliterated form), meaning ‘leaf’, Zebar is pronounced after the first phoneme, ‘b’. The word with all accents looks like “بَرگ”. The pattern for this word is CVCC. In figure 3.6, the first line shows the short vowel, Zebar, in red, where it should be, between the first consonant ب and the second consonant ر. The second line shows the word in proper cursive format, and the third line shows how the word is written in regular Persian, without short vowels, and cursive writing.



Figure 3.6: Short vowel Zebar in the word برگ

The second short vowel phoneme /e/, called **Zir**, is a small diagonal accent that appears below a letter (ِ). It sounds like the letter “e” in the word “bet”. In the word “زشت” /zeft/ (zesht in Roman transliterated form), meaning “ugly”, Zir is pronounced after the first phoneme, “sh”. The word with all accents looks like “زِشت”. The pattern for this word is CVCC.

And finally, the vowel phoneme **Pish** /o/, is a small accent similar to the comma in English. It appears on the top of a letter (ُ). An example with similar CVCC pattern for

this vowel is “تَرْد” /tord/ (tord in Roman transliterated form), with all accents “تُرْد”, meaning “crispy”.

3.3 Ambiguities caused by the semi-consonantal Persian alphabet

In chapter two, I explained the importance of vowel graphemes in English writing and visualized some consequences of a shortage in grapheme-phoneme correspondence. Obviously, we cannot apply all those numbers and consequences to Persian, and then measure the exact ambiguity of a semi-consonantal Persian alphabet. Earlier in this chapter, I mentioned some of the similarities and differences between the English and the Persian writing systems. Persian writing comes with its own difficulties and inconsistencies between graphemes and phonemes. Addressing all the possible problems of the Persian writing system is a much broader topic than the scope of this paper allows. Therefore, I have chosen to focus on some of the immediate difficulties that readers of Persian face while reading a document.

3.3.1 Words with three consonant graphemes

In the word level of this section, I am trying to illustrate the complexity of perceiving a written form of a word in Persian, where short vowels are usually omitted. Since the main focus of this section is on measuring the absence of short vowels, using an example that has long vowels in addition of short vowels makes it more difficult to demonstrate my point. Later, in chapter four, I will explain the role of long vowels in words, giving more details and examples.

I will use a very simple word, “گَرْد”, with certain meanings, to demonstrate and measure the possible level of ambiguity of this word. In Persian, گَرْد is composed of three consonant graphemes, گ, ر and د, as shown in table 3.4.

Persian Letter	Name of letter in Persian	Transliterated Name of Letter	English Equivalent Letter	IPA
گ	گاف	Gāf	G	/g/
ر	ر	Re	R	/r/
د	دال	Dāl	D	/d/

Table 3.4: Letters Gāf, Re and Dāl

As I stated earlier, short vowels are absent in the Persian writing system. The word گرد is written in a regular format of Persian writing, and as a result, the grapheme-phoneme correspondence ratio of this combination is not clear at this stage.

The other uncertainty created by the absence of short vowels in the example above is the exact meaning of the word. This group of letters may refer to three different words that have a similar visual representation in Persian orthography.

This takes us back to the example from Chapter 2 that focused on BAT with the vowel A removed from it. For the purposes of our thought experiment, I use the Persian word گرد.

Recall that in looking at a structure like 'bt', we did not have clear instructions for pronouncing the word, or for knowing where to add vowels or which vowels to use. Therefore, 'bt' lacked an immediate meaning.

We can perform a similar thought experiment using گرد.

Looking at a structure like گرد does not give us clear instructions for pronouncing the word, or for knowing where to add vowels or which what vowels to use. So again, there is not an immediate meaning to گرد.

Remember that Persian is read from right to left. The following Table 3.5 illustrates the formation of the Persian word with the equivalent English letters:

(←———— Reading Direction)

د	ر	گ	Persian Letters
D	R	G	English Equivalent
/d/	/r/	/g/	IPA

Table 3.5: The three consonants in the گَرِد combination

Here is the Persian word with potential places to add or skip short vowels shown by highlighting:

د _ ر _ گ (←)

Here is the English transliteration:

(→) G _ R _ D

Referring to the three short vowels that I mentioned earlier in section 3.2.3.2, we may potentially use either of those phonemes or choose not to use any vowel after each consonant. (I have not added any short vowels at the beginning or end of this example due to a different word-ending rule and exception, which will be explained in detail in sections 4.2.2 and 4.2.4). In total, we have four possibilities for adding short vowels after the first two letters, which is 4^2 possible formations in the گَرِد example, for a total of 16. Table 3.6 shows these possible combinations.

(← Table from right to left)

No.	Final Combination	د Dāl /d/	VS + ر + VS + گ	Second possible vowel insert	ر Re /r/	VS + گ	First possible vowel insert	گ Gāf /g/
1	گرد	د (d)	گرد	-	(r)	گ (gA)	- (A) /æ/	گ (g)
2	گرد	د	گرد	-				
3	گرد	د	گرد	-				
4	گرد	د	گرد	{ } or -				
5	گرد	د	گرد	-	ر	گ (gE)	- (E) /e/	گ
6	گرد	د	گرد	-				
7	گرد	د	گرد	-				
8	گرد	د	گرد	{ } or -				
9	گرد	د	گرد	-	ر	گ (gO)	- (O) /o/	گ
10	گرد	د	گرد	-				
11	گرد	د	گرد	-				
12	گرد	د	گرد	{ } or -				
13	گرد	د	گرد	-	ر	گ (g{})	- { }	گ
14	گرد	د	گرد	-				
15	گرد	د	گرد	-				
16	گرد	د	گرد	{ } or -				

{ } or -	Absence of vowel sound or consonant cluster
	Vowel Diacritics Zebar, Zir and Pish and consonant cluster
VS	Vowel Sound

Table 3.6: Possible letter combinations for the letters Gāf, Re and Dāl

In the 'گرد' (GRD) example, using the instructions I gave above, we can form 16 different combinations, of which three have meaning. In this experiment, only combinations without an intervening vowel in between the second and third consonants, which are Re and Dāl, are meaningful. We can also say that the second and third consonants in these

three words are a consonant cluster. As a result, these three examples have a CVCC pattern in which the V (the vowel sound) is one of the three short vowels.

At the beginning of this section, I mentioned the *Gāf*, *Re* and *Dāl* combination and possible word formations, but I did not touch on the meaning. The reason for this goes back to the culture of writing in the Persian alphabet.

According to the Academy of Persian Language and Literature (recall, as I explained in section 1.3, the status of the Persian writing system), diacritical marks for short vowels is an option choice, and the common and regular method of writing in Persian is to omit these marks. As a result, texts seldom include any diacritical marks other than diacritical points (which make distinctions between consonant letters, not accents). Therefore, a word (in this case گرد) never contains vowel diacritical marks other than in exceptional situations. Readers must guess the exact meaning of a word by evaluating it based on the rest of the words in the sentence or by recalling words with the same grapheme patterns from their memory. The situation may become even more complicated if the reader faces a single word, unfamiliar words or a complex sentence. These situations happen frequently, and the absence of short-vowel diacritical marks applies to all words, not just certain ones. Based on all these explanations, a group of consonant graphemes may refer to a single word, several words, or even different types of words including verbs, nouns, adjectives, etc.

Here are the words that have meaning formed in Table 3.6 from rows 4, 8 and 12:

- گرد
- گرد
- گرد

In Table 3.7, I show the English equivalents of the Persian words and their meanings:

(← Reading Direction)

Meaning	IPA	Word in English Transliteration	د	ر	Short vowel sound in IPA	Short vowel sound	گ	Word in Persian
Powder	/gærd/	G ARD	D	R	/æ/	A	G	گرد
Round	/gerd/	G ERD	D	R	/e/	E	G	گرد
Champion/ Athlete	/gord/	G ORD	D	R	/o/	O	G	گرد

Table 3.7: Meaningful words made with combination of Gāf, Re and Dāl and pronunciations

Considering the number of limitations that I have added to the experiment, the number of consonant graphemes and the simplicity of the words, the number of possibilities increases dramatically in a variety of other examples.

If the word has more letters, including consonants and long vowels, the number of possible new words increases. The situation becomes more complicated when a word has long vowels, combinations of long vowels, long vowels as consonant sounds, or loan words or transliterations of nouns and names from other alphabetical and non-alphabetical systems. Obviously not all of the combinations are meaningful, nor do all of the meaningful words/combinations come to mind immediately to support the meaning of the whole concept of written communication.

One important function of vowels, particularly short vowels, when writing in Persian using Perso-Arabic script, is to provide accuracy, to make perceiving and understanding the concept and the meaning of a written piece easier. When short vowels are absent, it is easy to misrepresent or misunderstand words and sentences.

3.3.2 Words with four consonant graphemes

In section 3.3.1, I explained and quantified the level of uncertainty in a written word composed of three consonants. In this section, I will explain a word with four consonant graphemes. Here, I will use the word “بَهْمَن” /bæh'mæn/, which means “avalanche”, a Persian name, and also name of the eleventh month in the Iranian calendar.² Bahman is composed of four consonant graphemes م, ب, ه and ن, and has a CVCCVC pattern. Both vowels in this example are short, therefore both are omitted in the written form.

In this example, I am following the same rules as in the example of the word with three consonants, with three short vowels, or with no vowels after each consonant. There are four possibilities after each consonant grapheme, and again no vowel at the beginning or at the end of the word. Therefore, there are four possibilities to consider after the first three letters of the word which are ه, ب and م. We have 4³ possible formations in the بهمَن example, for a total of 64.

3.3.3 Sentence level

For the sentence level of this section, I am using the Persian translation of the sentence I used in section 2.3 and replacing the English name with the Persian name ‘Bahman’ since we already measured uncertainty of this word in section 3.3.2. There are a few differences between the experiment for the English sentence and the experiment for the Persian sentence, which I mention here and explain in more depth later in this section. Here is the English sentence:

‘Tom lost his hat in the market.’

And the Persian translation:

‘بَهْمَن كُلاَهَش را در بازار گم كَرْد.’ (← Reading Direction)

² بهمن/Bahman is 30 days long. It begins in the second half of January and ends in February in the Gregorian calendar.

It reads as 'Bahman kolāhash rā dar bāzār gom kard'.

The first difference between the experiment in section 2.3 and this one is that I am dealing only with short vowels in this example. Therefore, I am removing only the short vowels from the Persian sentence. By removing the short vowel phonemes, we have the following sentence, which is in the normal style of writing with Persian scripts (writing in the style above is not seen frequently, other than in elementary school books, and is culturally considered lower literacy or a childish way of writing).

‘ بهمن کلاهش را در بازار گم کرد . ’ (← Reading Direction)

The sentence above is formed with seven words. Following, I will explain the situations for every word and then give the total number of possibilities for the sentence.

3.3.3.1 بهمن (Bahman)

Previously in section 3.3.2 we found a total of 64 possibilities for new words for this word.

3.3.3.2 کلاهش (kolāhash)

In this word, we have four consonants with the long vowel Alef in the middle of it. The pattern of this word is CVCVCVC, but two of these vowels are short and therefore omitted in the written form. Here is the highlighted word pattern CVCVCVC (Vs in blue are omitted in the written form. Therefore, we have just CCVCC in the written form of the word.

The long vowel Alef in this word ‘کلاهش’ is just read as /ɒ:/, similar to the letter U in ‘support’. Persian words do not begin with a group of consonant clusters, and there should be an intervening vowel between the first two consonants of the word, therefore there are three possible short vowels for the first letter ‘ک’. There is an Alef after the second letter ‘ل’. The third letter is an Alef, which is already a vowel sound. The fourth letter is He ‘ه’, and we have three short vowels and the possibility of a consonant cluster. Following the rules for word level in this section, I am not adding any possible letters at

the end of the word. There are three options for the first letter and four options for the fourth, a total of twelve possible readings of the word.

3.3.3.3 را (rā)

This word is formed with two letters, a consonant and a long vowel, the letters Re and Alef. It has a CV pattern. The letter Alef sits at the end of the word and works simply as a long vowel. Therefore, this word has only one form, and it does not create any additional ambiguity.

3.3.3.4 در (dar)

This is another word with two letters, but it has a CVC pattern. It is formed with the letters *Dāl* and *Re*, with a vowel in the middle. As mentioned in section 3.3.3.2, Persian words do not begin with a consonant cluster. Therefore, there are three possibilities available for this word.

3.3.3.5 بازار (bāzār)

This is an interesting example of a word where all the phonemes are written. The pattern for this word is CVCVC, and both vowels are long vowel Alef sounds /ɒ:/. It represents the exact verbal version of the word and it is not ambiguous at all.

3.3.3.6 گم (gom)

This word is similar to the word در (dar) in section 3.3.3.4. It has a total of three possibilities.

3.3.3.7 کرد (kard)

This word has three consonants in the written form, but a CVCC pattern. The vowel is short and therefore omitted in the written form. There are three possibilities for the first consonant grapheme and four possibilities for the second. So there is a total of twelve options for this word.

3.3.3.8 Overall number of possibilities for the sentence

To find the total number of possibilities for the sentence, I am removing the two words that have only one possible reading form, which are را (rā) and بازار (bāzār), from the calculation. The rest of the five words are in the Table 3.8 below:

(← Reading Direction)

Number of Possibilities	Word in English Transliteration	Word in Persian
64	Bahman	بهمن
12	kolāhash	کلاهش
3	dar	در
3	gom	گم
12	kard	کرد
94	Total	

Table 3.8: Number of possibilities for each word and the total number

There are 94 possible readings or pronunciations available for the sentence I used in this exercise. There are two difference between this example and the English version of the exercise. First of all, the situation in Persian writing is real and true, while the English example was designed simply to illustrate the situation. The second difference is that I removed all the vowels, both long and short, in the English sentence, while in the Persian example I removed just the short vowels.

3.4 Implications of having no short vowels in Persian orthography

In section 2, I visualized and illustrated what happens if we remove vowel graphemes from English orthography. In section 3, I used the same methodology as in section 2 for the Persian orthography, with some adjustments and differences. The main difference between these two experiments is that the Persian writing is a real case, while the English one is imaginary (used as a thought experiment). The main adjustment I have made in section 3 is to clarify the distinction between two groups of vowel graphemes, explained as short vowels and long vowels, in contrast with the case I made for English, which was about both kinds of vowels.

Quantifying the level of ambiguity of a word or a sentence without considering the entire context is not always the most accurate way of illustrating this situation. As I mentioned in sections 2.1.1, 2.1.2 and 2.3, words and sentences are parts of a concept or a meaning in a larger chain of words and sentences, which create whole communication. This fact is applicable to most languages. For example, the English word *read* could refer to more than one verb tense, but we have no clear idea which tense when we see it standing alone. Likewise, the Persian word *کی* has two distinct meanings, *when* and *who*. Therefore, context is an important factor for digesting and perceiving communicational material and its components, but clear and accurate components in a communicational system make the whole concept and point more certain, definite and understandable. The result is reduced ambiguity.

However, the increasing use of single words and short sentences in wayfinding, text messaging and user interfaces suggests an increasing the importance for these small components of communication.

In the next chapter, I will explain other uncertainties created by long vowels and long and short vowel combinations, and the consonant-vowel role of long vowels.

4. Complications Resulting from Vowel Combinations and Vowels Representing Both Vowels and Consonants (*Mater Lectionis*)

In this chapter, I will explain additional areas of confusion, ambiguity and possible misrepresentation in Persian orthography. Vowel combinations can appear in different forms in written Persian. There are cases when two or three of the long vowels combine together, or one of these long vowels combines with any of the short vowels.

Sometimes long vowels represent a consonant sound (this situation, called *Mater lectionis* will be explained in section 4.2), or other phonemes that are not represented by any letter.

Below, I address these situations in detail and explain how they compromise the clarity of communication in the Persian writing system.

4.1 *Alef*, a long vowel and a carrier

Alef is the only letter in Persian orthography that only represent vowel graphemes. *Alef* has several different roles in Persian writing. Table 4.1 shows the letter *Alef* in different positions.

Letter Position	Alef	Notes
Isolated	ﺀ	Alef is one of the seven letters that never connect to the letter that follows it. (See Section 3.2.2.) It may connect to the letter that precedes it, regardless of its position in the word.
Initial	ﺀ	Alef at the beginning of a word is usually a carrier of the short vowels Zebar, Zir and Pish.
Medial and Final with connection	ﺀ	Alef in the middle or at the end of a word may connect to the preceding letter if that letter is not one of the seven described in Section 3.2.2.
Isolated with diacritical mark	ﺀ̄	The only form of Alef that is certain to represent the long vowel /ɔ:/ usually comes at the beginning of a word, but it can also appear in other positions.

Table 4.1: Alef and its different contextual forms

As I explained in Section 3.2.3.1, Alef is the first long-vowel grapheme in the alphabet set. Alef can be a carrier for a short vowel when that vowel appears at the beginning of a word. When Alef appears at the beginning of a word, it stands alone, as I explained in Section 3.2.2. When it appears in the middle or at the end of a word, it may be connected to the preceding letter as long as that letter is not one of the seven mentioned in Section 3.2.2. If the preceding letter is one of those seven, then Alef must be independent. When Alef sits at the beginning of a word, it appears in one of two forms, “ﺀ̄” and “ﺀ”. The first form represents the long vowel Alef /ɔ:/; the second form is the carrier for the short vowels /æ/, /e/ and /o/.

Persian Word	Meaning	IPA	Roman
آتش	Fire	/b:tæʃ/	ātaš
اسب	Horse	/æsb/	asb
چاپ	Print	/tʃɒ:p/	chāp
ابزار	Tool	/æbzɒ:r/	abzār

Table 4.2: Four words showing the letter Alef in different roles

Table 4.2 shows four examples of the different roles the letter *Alef* can take. In the first row, Alef is the long vowel /b:/ at the beginning of the word; it appears as “آ”. The word has three letters and the VCVC pattern (the second vowel is short and omitted). In the second row, *Alef* sits at the beginning of the word, but now it is the placeholder for *Zabar*. Therefore, it appears as “ا”. (Whenever *Alef* sits at the beginning of a word as “ا”, it represents one of three short vowels.) This word has three letters too, but its pattern is VCC.

In the third row is a word with a CVC pattern in which the letter Alef represents the long vowel /b:/. There are two Alefs in the fourth row; the first one is the short vowel /æ/, and the second one is the long vowel Alef. This word has a VCCVC pattern.

The letter Alef can appear in the following positions as the first letter of a word:

4.1.1 Alef as the long vowel /b:/

Alef can begin a word as its principal role to represent the long vowel /b:/. In this situation, any letter after Alef will sound like a consonant, whether it is a regular consonant, or the letters Vāv or Ye, which are usually vowels. When Alef sits at the beginning of a word as /b:/, it appears in the form “آ” and does not pair with other vowels

to create a diphthong. All words that begin with “آ” start with the pattern VC... There are some exceptions: in loan words, where “آ” sits in the middle of a word; or in Persian compound nouns, where the second noun starts with “آ”, in which case it is written as a single noun. An example of Alef at the beginning of a word is “آرام” /ɑ:rɑ:m/, meaning ‘calm’, which has a VCVC pattern.

4.1.2 Alef with letter the Vāv as the long vowel /u/

The other two long vowels, Vāv and Ye, cannot sit alone at the beginning of a word to play long vowel roles (this is different from when they act as consonants, as will be described in Section 4.3), and they must pair up with an Alef (this will be explained in Section 4.2.1). An example of Alef and Vāv at the beginning of a word is “او” /u/, meaning ‘he/she’, which has a V pattern.

4.1.3 Alef with the letter Ye as the long vowel /i:/

The other combination is Alef and Ye, creating the sound /i:/ at the beginning of a word. An example of this is “ایل” /i:l/, meaning ‘tribe’, which has a VC pattern.

4.1.4 Alef standing alone as a placeholder for Zebar /æ/

As I said in Section 3.2.3.2, there is no letter dedicated to short vowels in the Persian writing system; short-vowel accents are usually omitted. Therefore, in order to start a word with a short vowel, there has to be an Alef to represent the absence of a short-vowel phoneme. The first short vowel is Zebar. In this case, Alef appears as “ا”. An example of Alef and Zebar at the beginning of a word is “انار” /ænɑ:r/, meaning ‘pomegranate’, which has a VCVC pattern.

4.1.5 Alef standing alone as a placeholder for Zir /e/

The second short vowel is Zir. Alef appears as “ا” at the beginning of a word to represent /o/. An example of Alef and Zir at the beginning of a word is “امروز” /emruz/, meaning ‘today’ which has a VCCVC pattern.

4.1.6 Alef standing alone as a placeholder for Pish /o/

Finally, the short vowel *Pish* at the beginning of a word is represented by an Alef. This combination appears in two forms, “ا” and “او”, and both represent /o/. In Section 4.3, I will explain the other possible instance of “او” at the beginning of a word. An example of Alef and Pish at the beginning of a word is “استان” /ostɑ:n/, meaning ‘state’, which has a VCCVC pattern.

Table 4.3 shows the six possibilities of using Alef at the beginning of a word.

Row	Persian Word	Meaning	IPA	Roman
1	آتش	Fire	/b:tæʃ/	ādash
2	اوستا	Master	/ustɑ:/	oustā
3	ایستا	Static	/i:stɑ:/	istā
4	اسب	Horse	/æsb/	asb
5	اسفند	Name of a month in the Persian calendar	/esfænd/	esfand
6	ارسی	Sash window	/orosi:/	orosi

Table 4.3: The six possibilities of using Alef at the beginning of a word

4.2 *Mater lectionis* in Persian

The other two vowels in Persian, Vāv and Ye, function as both long-vowel graphemes and consonant graphemes. This situation is called *Mater lectionis*, from the Latin *māter* + *lectiōnis*, literally ‘mother of the reading’. It refers to consonants that represent a specific vowel in writing systems like Hebrew and Arabic. The *Oxford Dictionary* defines this as “a consonantal sign, which indicates that a vowel sound is to be supplied” (*Oxford Dictionary*, Online). These two letters can represent consonant graphemes at any place in the word. But as vowel graphemes, they have to be written with an Alef when they appear at the beginning of a word. Table 4.4 shows the letter Ye in both a consonant and a vowel situation.

Persian Word	Casual Writing	Meaning	IPA	Roman
دیر	دیر	Late	/di:r/	dir
دیر	دیر	Monastery	/dejr/	deir

Table 4.4: The letter Ye as a consonant and as a vowel!

In Table 4.4, the first row shows Ye as a vowel, the second row shows it as a consonant. In both examples, Ye is the second letter of the word. I added all auxiliary accents in the second word to distinguish between the words. But in reality, and according to the Academy of Persian Language and Literature, both words are written in the “دیر” form. This means that the reader must guess what the word means by taking in the whole sentence. The pattern of the first word is CVC; that of the second is CVCC. But in Persian, both words have just three letters.

Persian Word	Casual Writing	Meaning	IPA	Roman
توپ	توپ	Ball	/tʊp/	tup
جوان	جوان	Young	/dʒævɒ:n/	javān

Table 4.5: The letter Vāv as a consonant and a vowel

Table 4.5 shows two examples of a similar situation for the letter Vāv. The top row shows a word that Vāv sits as a vowel grapheme; the second shows a word that Vāv sits as a consonant grapheme. In both examples, Vāv is the second letter. The pattern of the word in the first row is CVC; that of the word in the second row, CVCVC.

4.2.1 Vāv and Ye at the beginning of a word

Both Vāv and Ye can sit at the beginning of a word, but there are few situations where this happens. If either one sits at the beginning of a word as a single letter, it represents a consonant sound. But to represent the vowel functionality of either of these letters (/u/ and /i:/), we have to add an Alef in front of them.

Persian Word	Casual Writing	Meaning	IPA	Roman
وَرزِش	ورزش	Sport	/værzɛʃ/	varzesh
او	او	He/She	/u/	oo
یخ	یخ	Ice	/jæχ/	yakh
این	این	This	/i:n/	een

Table 4.6: The letter Vāv and Ye at the beginning of a word

In Table 4.6, the first two rows show the letter Vāv sitting at the beginning of the word. In the first row, Vāv is the first letter and a consonant grapheme. In the second row, it

sits in second place to represent a vowel grapheme with the support of Alef. Rows three and four show the same pattern using the letter Ye.

Alef in rows two and four is an auxiliary letter. We can consider Alef + Vāv and Alef + Ye as digraphs. But when these combinations sit at the beginning of a word, Alef plays the role of a vowel grapheme, and the combination changes to VC, or a diphthong.

4.2.2 Vāv and Ye at the end of a word

Unlike the situations above, when Vāv and Ye occur at the beginning of a word, when they sit at the end of a word they can function as either a consonant or a vowel without combining with another letter.

Persian Word	Casual Writing	Meaning	IPA	Roman
سَرَو	سرو	Cypress	/særv/	sarv
بو	بو	Smell	/bu/	bou
پِی	پی	Foundation	/pej/	pey
بازی	بازی	Game	/bɑ:zi:/	bāzi

Table 4.7: The letters Vāv and Ye at the end of a word

In Table 4.7, in the two first rows Vāv sits at the end of the word. In the first row, Vāv is a consonant grapheme, and in the second row a vowel grapheme. Rows three and four follow the same pattern using the letter Ye. In all these examples, Vāv and Ye are functioning in their respective roles alone, without any supporting letter.

4.2.3 Vāv and Ye in the middle of a word

Finally, either of these letters can sit in the middle of a word and function as either a vowel grapheme or a consonant grapheme.

4.2.4 Two exceptions

As I mentioned in Section 3.3, there are many examples and situations where these rules are not applicable. And there are many additional rules. Two of these cases are more immediate when we are talking about short and long vowels.

The letter Vāv is one of the most ambiguous in the Persian alphabet. It can play the role of a short vowel sound in some words. For example, in the word “رویا”, /rojɒ:/, meaning ‘dream’, Vāv, which is the second letter, represents the short vowel /o/. But in the word “گویا” /gujɒ:/, meaning ‘communicative’, Vāv represents the long vowel /u/. Not all /o/ phonemes are represented by Vāv. Therefore, the letter can play three different roles: the long vowel /u/, the short vowel /o/, and the consonant /v/. There is a fourth sound, /ow/, which I will explain in Section 4.4.

As I described in Section 3.2.3.2, the short vowel phoneme Zir usually has no assigned grapheme or visual representation. But there is an exception. When a word ends with the sound /e/, the letter **He** (/he/), with an isolated form of “ه”, can play the role of the short vowel /e/. But *He* is not always a short vowel Zir when appears at the end of words, it may also function as its original consonant grapheme. This situation happens only at the end of a word, not at the beginning or in the middle, or when the preceding letter is not Alef. An example of this situation is “باده” /bɒ:de/, meaning ‘wine’. Here the word pattern is CVCV. In other cases, the letter He sound is /h/, like the English letter H, whether it comes at the beginning of a word, in the middle, or at the end when it is not representing Zir.

4.3 Vowel combinations and digraphs

Vowel combinations in Persian orthography is one of the areas of most ambiguity in the language. The location of vowels is very important here as well. As I said in section 3.2.3.1, written Persian includes three long vowels. Digraphs in Persian may be composed of one of the following combinations:

1. Consonant + Consonant
2. Consonant + Vowel
3. Vowel + Consonant
4. Vowel + Vowel

Each of these situations may appear in words, and the reader has to figure out the right form of the word by referring to memory or by understanding the entire meaning of the material. In combination 1, both letters are acting as a consonant sound, joining together with a short vowel in the middle (a CVC pattern) or as a consonant cluster in a larger word. For example, in the word 'یواش' /jævɒ:ʃ/ (yavash in its Roman transliterated form), meaning 'slowly', the first two letters, Yā and Vāv, act as consonants joined by the short vowel *Zabar*, followed by an *Alef*.

An example of combination 2 is the word 'ویران' /vi:rɒ:n/ (virān in Roman transliterated form), meaning 'ruined'. The letter Vāv at the beginning of the word represents the consonant V /v/, followed by Yā as the long vowel /i:/.

The word 'گویش' /gujeʃ/ (guyesh in its Roman transliterated form), meaning 'dialect', is an example of combination 3. Here the letter Vāv acts as the long vowel /u/, and the letter Yā acts as the consonant /j/.

And finally, in combination 4, vowel + vowel, usually an Alef joins either Vāv or Yā. In the word 'ایران' /i:rɒ:n/, for example, the letter Alef at the beginning is followed by the letter Yā, to represent the long vowel /i:/ in the word Iran.

Sometimes these digraphs represent a monophthong and sometimes a diphthong. In these situations, the reader is required not only to guess which possible short vowel to add, but to understand whether these letters represent a consonant sound or a vowel sound, and then to distinguish between different possible words. For example, the word اوستا represents two words, Avesta and Ousta. The first one combines Vowel + Consonant at the beginning. It means the religious book of Zoroastrianism. The second combines Vowel + Vowel at the beginning and it means master.

4.4 Unrepresented and unrecognized phonemes or diphthongs

The Academy of Persian and education system in Iran categorized the vowel sounds of the Persian language in two groups of long vowels and short vowels each with three members. As a result, some other phonemes of the Persian language are unrecognized. These phonemes are part of the Tehrani dialect and some other dialects of Persian, inside and outside Iran. One example of these phonemes is /ow/, which represent by letter Vāv. Letter Vāv mapped for two phones, /o/ and /u/. But both society and officials are making distinguish between all these three different phones in Roman transliterated Persian. In the transliterated versions letter O adopted for short vowel /o/, one of OO, OU or U adopted for long vowel /u/ and finally OW adopted for /ow/.

In 2012, Iranian government proposed a “New Persian Romanization System” to the Tenth United Nations Conference on the Standardization of Geographical Names, in which vowels of Persian language counted as eight, in contrast to what Persian sources are introducing. One of these eight phonemes is /ow/. (p. 5, UN)

5. Discussion

The majority of modern language writing systems have some inconsistencies and shortages in relation to their spoken version. Some of these shortages are due to the nature of verbal languages, which are more flexible and can be aided by gestures, tones, and other non-verbal tools of communication. The Persian writing system has its own additional difficulties. Persian is adapted from Arabic, which is not the best fit for the language (as was explained in Section 1.3). Also Persian missed the optimization that happened to Roman scripts as far as the introduction of vowels through the Greeks' adaptation of the Phoenician alphabet.

This doesn't mean that through adaptations and optimizations, all inconsistencies and shortcomings are resolved. And it is true that there are variety of inconsistencies between languages that use the Roman alphabet. For example, Roman scripts are not perfectly optimized for English. We can see this in the large difference between the number of English phonemes compared to the number of English letters. It can be measured by the phoneme-grapheme correspondence of English orthography (for a more detailed explanation, see Chapter 2). This caused some reduction of optimization; such as the disappearance of the thorn or þorn (þ), which was replaced by the digraph *th*.

In this research I introduced an analytic approach as a tool to examine the difficulties caused by the lack of short vowel letters and the inconsistency of long vowel letters. I visualized the consequences of having no vowels in English at word level and at

sentence level to arrive at a quantified overview of the function of short vowels. Then I applied the same methodology, with a few necessary adjustments, to Persian. This experiment gave me analytical data to support the idea that the Persian writing system is not well-suited to the Persian language.

A possible critique of this idea is that the context of words in a sentence mitigates or eliminates confusion in the visual representation of language. It is true that context is what makes reading possible; a single word is often meaningless. However, this does not negate the fact that the lack of representation of certain sounds in the Persian script is a problem that leads to more ambiguity than is found in scripts that represent, however imperfectly, all the sounds needed to fully form a word. This tool focuses on the components of written communication and on how the lack of representation of spoken components may cause misunderstanding of the whole meaning.

This analytical process focuses on the ambiguity inherent in the script at the level of the word, not a larger amount of text. Because we are in the age of short, sometimes 'word length' communication like text messaging, wayfinding and maps, this potential ambiguity is more likely to inhibit communication. Finally, this tool is necessarily designed for a particular group of measurements, and not intended as a complete analysis of how successfully Persian script represents language.

This analytic approach could be used as a foundation for future investigations into different levels of ambiguity in many alphabetical writing systems, with considerations and modifications appropriate to those systems.

Persian is a complex and ambiguous writing system. It is beyond the scope of this paper to design or propose a new writing system. Furthermore, writing systems are notoriously difficult to alter, upgrade or improve.

Proposing some minor reforms, including introducing of new letters to replace missing short vowels, is unlikely to solve this particular problem; instead, it would most likely simply shift some of the current difficulties to other areas. During the process of this

research, I spent some time trying to discover possible new short vowels for the current Persian writing. But there are two major problems:

- proposing new letters would require changing the entire Persian orthography,
- before we design a new letter, we have to face other problems in Persian scripts, such as cursive writing, which requires considering the variety of contextual forms of a letter, RTL writing order in contrast with LTR scientific notation ...

Therefore, I preferred not to include any of my trials.

Some of the problems of Persian script are particularly significant in the modern world of telegraphic communication such as texting and wayfinding environments. In these situations, a word or a short phrase represents a larger body of information, and there is no further supporting information that can provide contextual clues that would reduce confusion or uncertainty. Also, though it is not a focus of this paper, small diacritics in Persian are likely to be difficult to read for aging populations. Ironically, the computer, which enabled ligatures to be automatically composed, may have acted as a brake to changes that might have naturally evolved otherwise.

Transitioning from one writing system to another would be no small matter. It would require careful consideration, planning and investment, and would need to be informed by a strong base of evidence. There are cases of writing systems that have changed, most notably simplified Chinese in the mid-20th century. Some alterations have been minor, such as German spelling reforms; others have been radical, such as the Turkish reforms of Kemal Ataturk in the early 20th century, when Arabic was replaced with Roman.

When we add other social considerations, such as the problem of cultural resistance, particularly resistance to the hegemony of Western culture, it becomes evident that the widespread Romanization of Persian is neither practical nor desirable. The Chinese approach of modifying the script while simultaneously maintaining its essential nature might be the most advantageous.

Clearly, Persian-speaking societies would benefit from finding a solution for improving the orthography and morphology of their written language as soon as possible.

The Persian writing system, its characteristics and situation, is a subject of disagreements between linguists. One group of linguists wants a fundamental changeover of the writing system, while others oppose the idea. Both schools of thought offer supporting evidence. One linguist in favour of introducing a new writing system is Mohammad-Reza Bateni. In an interview with Alinejad and Roshan for the BBC Persian (BBC Persian, 2008), Bateni argues that the current writing system has no potential for improvement, and that the Persian language needs a new writing system that is developed entirely for itself. Otherwise the problems of the writing system will force us to take action. But even Bateni believes that in the short term, we need at least a grapheme for the short vowel Zir in the alphabet set.

In my opinion, regardless of what these school of thought believe, the society of Persian speakers is suffering because of the current writing system's problems. We may postpone the decision to keep the current system or developing an entirely new one, but in the short term we need to make some adjustments and modifications. These adjustments may include redesigning some letters or introducing missing vowel sounds.

6. Conclusion

Writing is the most common and culturally dependent form of non-verbal communication. Regardless of the context of the communication, scripts of the writing system are the medium between the language of the written document and the linguistic understanding of the reader. Verbal communication on its own has many layers of elements, including tones, accents, dialects, environmental elements and gestures that are absent in written communication. These additional layers of information are usually not part of written communication; they are expressed only in a rudimentary way by the use of grammatical structure, punctuation marks and typographical treatments. These elements address a fundamental lack of written communication, but they do not fully replace the metacommunication of oral language. Therefore, a well-coordinated writing system can increase accuracy and reduce ambiguity of written materials.

In this study, I tried to illustrate and demonstrate one of the existing shortages of the Persian writing system, which has an incomplete set of vowel sounds to represent phonemes of the Persian language. Millions of people use this writing system every day; at the same time, the popularity of digital platforms creates vast amounts of written materials. The increasing literacy rate and the influence of technology make it necessary to reassess the existing writing system in order to address the needs of the society.

There are two main schools of thought regarding the Persian writing system;

- The first suggests keeping the existing writing system. Its main argument is that we need to save the Persian language and cultural identity by avoiding non-Persian influences. However, minor reforms of the writing system are considered acceptable.
- The second school goes further. It considers the current script system an incorrigible platform for the Persian language and suggests changing it fundamentally. Most scholars in this group support the idea of romanizing the Persian writing system.

Regardless of what these schools of thought advocate, we cannot fundamentally change or solve all the problems of a writing system by modifying some aspects of it. But the current status of the Persian language and writing system requires taking action. Leaving them as they are will cause many problems for its users, and these problems are growing. Introducing a new writing system requires long-term planning along with investing in areas such as education.

BIBLIOGRAPHY

- Adobe Inc. (n.d.). Glossary of typographic terms. Retrieved November 7, 2018, from <https://www.adobe.com/ca/products/type/adobe-type-references-tips/glossary.html>
- Alavi, B. (2013). *Varaq Pāreh'hāye Zendān* [Scrap Papers from Prison] (3rd ed.). Tehran: Negāh.
- Alinejad, S., & Roshan, S. (2008, June 16). مصاحبه با دکتر باطنی؛ خط فارسی را نمی توان اصلاح کرد. Retrieved December 14, 2018, from http://www.bbc.com/persian/arts/story/2008/06/080616_an-cy-bateni-interview.shtml
- Bailey, P. G., Tuinman, J. J., & Jones, S. (2013, December 16). Literacy. *The Canadian Encyclopedia*. Retrieved May 30, 2016, from <https://www.thecanadianencyclopedia.ca/en/article/literacy>
- Brecht, B. (1978). *Brecht on Theatre: The development of an aesthetic* (J. Willett, Ed.; J. Willett, Trans.). London: Methuen Drama.
- Bringhurst, R. (2005). *The solid form of language: An essay on writing and meaning*. Kentville, N.S.: Gaspereau Press.
- Carnegie Mellon University. (2014, November 19). The CMU Pronouncing Dictionary. Retrieved May 8, 2018, from <http://www.speech.cs.cmu.edu/cgi-bin/cmudict#phones>
- Carr, P. (2008). *A glossary of phonology*. Edinburgh: Edinburgh University Press.
- Daniels, P. T., & Bright, W. (Eds.). (1999). *The World's Writing Systems*. London: Oxford University Press.
- diacritic. 2018. In *Merriam-Webster.com*. Retrieved September 28, 2018, from <https://www.merriam-webster.com/dictionary/diacritic>
- Doward, J. (2016, May 29). Boys who live with books 'earn more as adults'. *The Guardian*. Retrieved May 30, 2016, from <https://www.theguardian.com/education/2016/may/29/boys-books-earnings-adults>
- Fengxiang, F., & Altmann, G. (2008). Graphemic representation of English phonemes. *Quantitative Linguistics Analyses of Script*, (63). doi:10.1515/9783110207101.2.25
- Functional Illiterate. 2016. In *Merriam-Webster.com*. Retrieved May 30, 2016, from <https://www.merriam-webster.com/dictionary/functionalliterate>

- Drucker, J. (1999). *The alphabetic labyrinth: The letters in history and imagination*. London: Thames and Hudson.
- Free Vector Maps. (2017, May 30). Map of World with Countries - Multicolor [Map]. In *Free Vector Maps*. Retrieved December 17, 2018, from <http://freevectormaps.com>
- Gelb, I. J. (1989). *A study of writing*. Chicago u.a.: University of Chicago Press.
- Hammer, K. (2018, May 08). Global rate of adult literacy: 84 per cent, but 775 million people still can't read. Retrieved September 24, 2018, from <https://www.theglobeandmail.com/news/world/global-rate-of-adult-literacy-84-per-cent-but-775-million-people-still-cant-read/article4528932/>
- Hashabeiky, F. (2005). *Persian Orthography, Modification or Changeover? (1850–2000)* (Doctoral dissertation, Department of Linguistics and Philology, Uppsala University, 2005). Stockholm: Elanders Gotab.
- Hyman, L. M. (1975). *Phonology: Theory and analysis*. New York: Holt Rinehart & Winston.
- Indicators of Well-being in Canada. (n.d.). Learning - Adult Literacy. *Well Being*. 30 May 2016. Retrieved May 23, 2016, from <http://well-being.esdc.gc.ca/misme-iowb/.3ndic.1t.4r@-eng.jsp?iid=31>
- IPA. (n.d.). Retrieved December 15, 2018, from <https://www.internationalphoneticassociation.org/>
- IPA charts and sub-charts in four fonts. (n.d.). Retrieved December 15, 2018, from https://linguistics.ucla.edu/people/keating/IPA/IPA_charts_2018.html
- IPA. (2018). THE INTERNATIONAL PHONETIC ALPHABET (revised to 2018) [PDF file]. Retrieved from https://linguistics.ucla.edu/people/keating/IPA/IPA_DejaVu_2018_full.pdf
- Iranian Committee for standardization of Geographical Names. (2012, July 31). *New Persian Romanization System (Proposed for adoption)* [PDF file]. New York: Tenth United Nations Conference on the Standardization of Geographical Names.
- Jackson, D. (1981). *The Story of Writing*. New York, New York: Taplinger Publishing.
- Academy of Persian Language and Literature. (n.d.). *Dastoor-e Khatte Farsi* [Persian Writing Guideline] (13th ed.). Tehran: Academy of Persian Language and Literature.
- Academy of Persian Language and Literature. (n.d.). درباره فرهنگستان زبان و ادب فارسی [About APLL]. Retrieved November 14, 2018, from <http://apll.ir/درباره/>
- mater_lectionis. 2017. In [oxforddictionaries.com](https://en.oxforddictionaries.com/definition/mater_lectionis). Retrieved April 7, 2017, from https://en.oxforddictionaries.com/definition/mater_lectionis
- National science education standards* (Second ed.). (1996). Washington, DC: National Academy Press.
- Ostadi, K. (2012). تاریخچه اندیشه تغییر خط در ایران [History of the Thought of Changing Script in Iran]. *Ayene-ye Pazhoohesh*, 23(134), 22-38. Retrieved October 21, 2018, from http://jap.isca.ac.ir/issue_58_74_Volume_23_Issue_134_Autumn_2015.html

- Pflughaupt, L. (2008). *Letter by letter: An alphabetical miscellany*. New York: Princeton Architectural.
- Prods Oktor Skjærvø, "IRAN vi. IRANIAN LANGUAGES AND SCRIPTS (3) Writing Systems," *Encyclopaedia Iranica*, XIII/4, pp. 366-370, available online at <http://www.iranicaonline.org/articles/iran-vi3-writing-systems> (accessed on 30 December 2012).
- Roach, P. (2011). *Glossary – A Little Encyclopaedia of Phonetics*. peterroach.net.
- Robinson, A. (2007). *The Story of Writing: Alphabets, Hieroglyphs & Pictograms* (2nd ed.). London: Thames and Hudson.
- Sadeghi, A., & Zandi-Moghaddam, Z. (2016). فرهنگ املائی خط فارسی [Persian Orthographic Dictionary] (7th ed.). Tehran: Academy of Persian Language and Literature.
- The World Bank. (2016). *Literacy rate, adult total (% of people ages 15 and above)*. Retrieved May 30, 2016, from <https://data.worldbank.org/indicator/SE.ADT.LITR.ZS>.
- UNESCO (1996). *Learning: The treasure within - Report to UNESCO of the International Commission on Education for the Twenty-first Century*. Paris. Available from http://www.unesco.org/education/pdf/15_62.pdf
- UNESCO (1996). *The Hamburg Declaration on Adult Learning. Hamburg*. Available from <http://www.unesco.org/education/uie/confintea/declaeng.htm>. Accessed 14 November 2017
- UNESCO (2015). *Mobile Phones & Literacy - Empowerment in Women's Hands*. Paris. Available from <http://unesdoc.unesco.org/images/0023/002343/234325E.pdf>
- UNESCO Institute for Statistics. (2013, June). *Adult and Youth Literacy National, Regional and Global Trends, 1985-2015* [PDF file]. Montreal: UNESCO.

APPENDIX A: The IPA Chart

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2018)

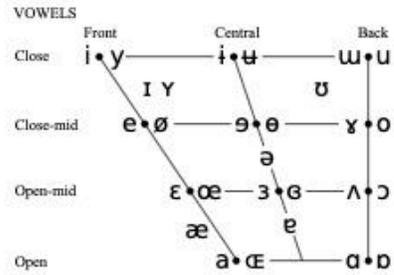
CONSONANTS (PULMONIC) © 2018 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			ʀ					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
◌ Bilabial	ɓ Bilabial	Examples:
◌ Dental	ɗ Dental/alveolar	◌ Bilabial
◌ (Post)alveolar	ɟ Palatal	◌ Dental/alveolar
◌ Palatoalveolar	ɠ Velar	◌ Velar
◌ Alveolar lateral	ɣ Uvular	◌ Alveolar fricative



OTHER SYMBOLS

◌ Voiceless labial-velar fricative	◌ Alveolo-palatal fricatives
◌ Voiced labial-velar approximant	◌ Voiced alveolar lateral flap
◌ Voiced labial-palatal approximant	◌ Simultaneous ʃ and x
◌ Voiceless epiglottal fricative	Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.
◌ Voiced epiglottal fricative	ts̺ kp̺
◌ Epiglottal plosive	

DIACRITICS Some diacritics may be placed above a symbol with a descender, e.g. ɲ̥̊

◌ Voiceless	◌ Breathily voiced	◌ Dental
◌ Voiced	◌ Creaky voiced	◌ Apical
◌ Aspirated	◌ Linguolabial	◌ Laminal
◌ More rounded	◌ Labialized	◌ Nasalized
◌ Less rounded	◌ Palatalized	◌ Nasal release
◌ Advanced	◌ Velarized	◌ Lateral release
◌ Retracted	◌ Pharyngealized	◌ No audible release
◌ Centralized	◌ Velarized or pharyngealized	
◌ Mid-centralized	◌ Raised	
◌ Syllabic	◌ Lowered	
◌ Non-syllabic	◌ Advanced Tongue Root	
◌ Rhoticity	◌ Retracted Tongue Root	

SUPRASEGMENTALS

- ◌ Primary stress ˈfoʊnəˈtʃən
- ◌ Secondary stress ˌ
- ◌ Long eː
- ◌ Half-long eˑ
- ◌ Extra-short ɛ̋
- ◌ Minor (foot) group |
- ◌ Major (intonation) group ||
- ◌ Syllable break .ji.ækt
- ◌ Linking (absence of a break) ~

TONES AND WORD ACCENTS

LEVEL	CONTOUR
◌ Extra high	◌ Rising
◌ High	◌ Falling
◌ Mid	◌ High rising
◌ Low	◌ Low rising
◌ Extra low	◌ Rising-falling
◌ Downstep	◌ Global rise
◌ Upstep	◌ Global fall

(Source: IPA Website, IPA charts and sub-charts in four fonts)