In the Name of God

دانشگاه شیراز

Shiraz University

Faculty of Literature and Humanities

M.A. Thesis in Teaching Persian to Speakers of Other Languages

A COMPREHENSIVE SCHEME FOR PHONEMIC TRANSCRIPTION OF PERSIAN IN THE ENGLISH ALPHABET

By

NAVID BARADARAN HEMMATI

Supervised by

Firuz Sadighi, PhD.
Mortaza Yamini, PhD.

June 2009

دانشکده‌ی ادبیات و علوم انسانی

پایان‌نامه‌ی کارشناسی ارشد در رشته‌ی آموزش زبان فارسی به غیرفارسی‌زبانان

الگویی جامع برای واج‌نویسی فارسی با الفبای انگلیسی

به کوشش

نوید برادران همتی

استاد راهنما

دکتر فیروز صدیقی
دکتر مرتضی یمینی

تیر ماه ۱۳۸۸

Declaration

I—Navid Baradaran Hemmati (850256)—a student of Teaching Persian to Speakers of Other Languages at the faculty of Literature and Humanities, hereby declare that this thesis is the result of my own research, and that I have provided exact references wherever I have quoted someone. I also declare that the topic of my research is an original one and has not been worked upon before. I hereby promise not to publish the findings of this research and not to make it accessible to others without the permission of Shiraz University. Shiraz University holds the copyright on this research.

N. Baradaran Hemmati

July 8, 2009

IN THE NAME OF GOD

A COMPREHENSIVE SCHEME FOR PHONEMIC TRANSCRIPTION OF PERSIAN IN THE ENGLISH ALPHABET
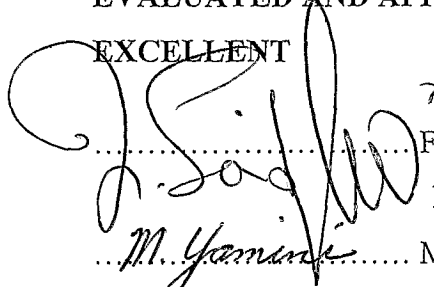
BY

NAVID BARADARAN HEMMATI

THESIS

SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS (M.A.)

IN

TEACHING PERSIAN TO SPEAKERS OF OTHER LANGUAGES

SHIRAZ UNIVERSITY

SHIRAZ

ISLAMIC REPUBLIC OF IRAN

EVALUATED AND APPROVED BY THE THESIS COMMITTEE AS:

EXCELLENT

.... F. SADIGHI, PhD.., PROFFESSOR OF LINGUISTICS (CHAIRMAN)

.... M. YAMINI, PhD., ASSISTANT PROFFESSOR OF ENGLISH LANGUAGE TEACHING (CHAIRMAN)

.... J. RAHIMIAN, PhD., ASSOCIATE PROFFESSOR OF LINGUISTICS

.... S.M. ASSI, PhD., ASSOCIATE PROFFESSOR OF LINGUISTICS

JUNE 2009

ACKNOWLEDGEMENT

I hereby thank everyone who has helped me develop the present thesis.

ABSTRACT

A COMPREHENSIVE SCHEME FOR PHONEMIC TRANSCRIPTION OF PERSIAN IN THE ENGLISH ALPHABET

BY

NAVID BARADARAN HEMMATI

In conveying Persian proper names or texts in contexts where the original alphabet is undesired, disallowed, or unavailable for non-linguistic reasons, and in the initial stages of Persian language teaching, an alternative or obligation is to transcribe Persian phonemes in the English alphabet. In order to standardize phonemic transcription of Persian in the English alphabet, a comprehensive scheme was to be developed. Besides its comprehensiveness, the scheme was decided to be scientific, simple, and popular. Based on the necessity of these features, suggestions made in the earlier similar schemes were discussed. The discussions led to the following conclusions: there are twenty-nine phonemes in the Persian sound system, and they should be represented by the symbols introduced; an /ʔ/ sound initiating a sequence or surrounded by two vowels should not be represented; phonological processes typical of spoken language, such as elision and metathesis, should be ignored; a longer vowel should be represented by the corresponding symbol followed by a colon; a geminated consonant should be represented by doubling the corresponding symbol; sounds affected by the phonological process of assimilation should be represented based on how they are pronounced in practice; the sequence /ɪj/ should always be represented as "ɪy"; the sequence [ow] should be represented as "ow"; linguistic phenomena other than phonological ones, whether morphological, lexicological, or syntactic, should be ignored; in the case of orthographic phenomena, spelling rules of the somehow related orthographies, as well as the original representations of Roman-based abbreviations should be ignored, and the word-division and punctuation systems, as well as the representations of numerals, symbols, and typefaces should be borrowed from British English. It is suggested that similar projects be performed on other languages, dialects, and accents.

V

TABLE OF CONTENTS

Title   **Page**

Title	Page

LIST OF TABLES

X

CHAPTER ONE

INTRODUCTION

1-0- Introduction

The Present chapter is an introduction to what the thesis contains. The chapter includes four sections. In the first section, the issue with which the study is concerned is clarified and the significance of the study is explained. In the second section, the objectives of the study are stated. In the third section, the research questions are listed, and finally, in the fourth section, the technical terms frequently used throughout the thesis are defined.

1-1- Statement of the Problem and Rationale for the Study

The Persian alphabet uses an Iranian version of Arabic script, with which many people in the world are unfamiliar. In order for them to be able to read Persian, the most straightforward way seems to be to learn the alphabet. Clearly, if a small text is concerned, this would not be the most economical way. To convey short, non-translatable Persian proper names internationally, for instance, one is to write them in the same language but in a more popular script. Besides, in certain cases, it is a matter of obligation rather than preference to use the English alphabet or, at least, another Roman-based alphabet. Instances include international official documents such as passports, international certificates, and names of

organizations and corporations.

Furthermore, there are cases where conveying pieces of language in the original alphabet is actually not practical. In exchanging typed texts electronically or via telecommunication (in services such as e-mail, SMS and chat, for instance), it is impossible to use a script unless it is technically recognized. Roman script is, in most cases, because it is the script upon which an interlingua as significant as English, as well as the languages of many developed western countries, are based. To type texts that ordinarily make use of non-recognized scripts, one has to switch to recognized ones such as Roman.

Moreover, it has been argued that even in the case of teaching Persian as a foreign language, "for the convenience of the users" (Yarmohammadi and Sadighi, 2005: Preface), one had better not begin with the original alphabet. This idea has been held and practiced in some Persian language teaching textbooks, such as Yarmohammadi and Sadighi (2005). In such works, Roman-based alphabets have been preferred. This seems to make perfect sense, as Roman script is globally widespread, and it is utilized by an interlingua as significant as English.

1-1-1- Transcription versus Transliteration

In all the above cases, the writer is involved in a process referred to as Romanization, "The use of the Latin alphabet to transcribe non-Latin writing systems." (Crystal, 1987: 429) There are two distinct approaches to Romanization: transliteration and transcription (These are terms with broader senses than just applying to Roman script, though). In the first approach, "each character of the source language is converted into a character of the target language" (Crystal, 1987: 346), whereas in the latter, "the *sounds* of the source words are conveyed by letters in the target

2

language." (Crystal, 1987: 346) "This [(that is, the latter)] approach is often unavoidable with languages that use partial alphabetic scripts (e.g. Arabic, p. 202), where transliteration would be very difficult" (Crystal, 1987: 346). Being a consonantal alphabet, "where the marking of vowels (using diacritics) is optional" (Crystal, 1987: 202), the Persian alphabet is among those that have to be transcribed rather than transliterated.

1-1-2- Phonemic versus Phonetic Transcription

Transcription is itself of two types: narrow and broad.

> 1 a **narrow transcription** (also called a **phonetic transcription**) is one that shows phonetic details (for example, aspiration, length, etc.), by using a wide variety of symbols and, in many cases, DIACRITICs. A **systematic phonetic transcription** shows all the phonetic details that can be recorded.
> 2 a **broad transcription** (also called an **impressionistic transcription**) is one that uses a simple set of symbols and does not show a great deal of phonetic detail. A **phonemic transcription** is a broad transcription that shows all and only those sounds that are distinctive PHONEMEs in the language being transcribed. (Richards and Schmidt, 2002: 560-1)

To satisfy the needs stated above, providing details beyond phonemes would be unnecessary, because it leads to more confusion than simplicity on the part of the reader, while simplicity is among the primary goals expected to be achieved through a device of this kind. Phonemic transcription, therefore, is the most suitable method here.

3

1-1-3- Need for a Scheme

Generally speaking, the development of standards is necessary where different individuals are required to behave in the same way on a certain occasion. Where no standards are recognized, it would be a matter of taste how one decides to act. To harmonize, you need to standardize. This particular case of transcription is not an exception. Without a standard, everybody does it in his/her own way. To standardize the application of phonemic transcription to Persian, there has to be an internationally agreed-upon scheme. "In the absence of an internationally agreed scheme, it is often very difficult to trace terms and names...." (Crystal, 1987: 346) What follows is a description of only part of the present chaotic situation:

- To transcribe the sound /u/, some use the letter "o": "zod" for /zud/ (early). This is probably because similar symbols are (and sometimes the same symbol is) used in the original alphabet for both /o/ and /u/ sounds.

- For the same sound, some others use the sequence "oo": "zood" for /zud/. Those might have been inspired by the spellings of English words such as "tool".

- Yet another group uses the sequence "ou" for the same sound. This might be a remnant of probable initial transcriptions by the early French residents of Persian-speaking countries, following the spellings of French words such as "vous" (/vu/, you).

- To transcribe the sound /i/, some use the letter "e": "der" for /dir/ (late). This might be due to the pronunciation of the name of the English letter "e".

- Persian speakers living in countries where Roman script is used sometimes seem to be following the spellings common in the languages spoken around them. Instances include the selection of the

4

letter "s" to represent the sound /z/, and the sequence "ch" to represent the sound /ʃ/ by Persian speakers living in Germany and France, respectively.

— The absence of an agreed-upon transcription scheme even seems to account for certain commercial abuses. Selection of the letter "s" to represent the last sound in transcription of the name of the Iranian product labeled "Panberes", pronounced /pamberiz/, might be explainable in terms of an effort to make the name resemble that of the popular similar product known as "Pampers" (This is only a pessimistic guess, though).

So long as no standards are established, however, none of the above, not even the abuser, can be blamed. This is just like a state of anarchy, where no one can be accused of breaking the law, because there is actually no law. The prescription of a unique transcription scheme would mean, as Guitarte and Quintero (1974: 349) put it, "putting an end to *orthographic anarchy*".

1-1-4- Comprehensiveness of the Scheme

A transcription scheme has to be comprehensive, in that it contains conventions not only for the citation form, "the pronunciation given to a word when it is produced in isolation, and not in CONNECTED SPEECH" (Crystal, 2003: 73), but also for pieces of connected speech. "Transcription is not only used to represent words in isolation but can also be employed for whole stretches of speech. In all languages, the pronunciation of words in isolation is very different from the way they appear in connected speech...." (Collins and Mees, 2003: 15) Devising all these conventions will make the scheme able of answering the variety of questions the transcriber may encounter. The more questions a transcription scheme leaves unanswered, the more question marks hang over its comprehensiveness (See 1-3 for

some of those questions).

1-1-5- Priority of the English Alphabet

Being used to represent the most significant interlingua, the English alphabet is perhaps the largest set of characters the validity of which is guaranteed in all international areas. Even other Roman-based alphabets sometimes have to replace certain graphemes of theirs (mainly diacritical ones) with alternative symbols so that they are conveyed as desired to. German "ä", "ö", "ß", and "ü", for instance, are quite frequently required to be replaced by the symbols "ae", "oe", "ss", and "ue", respectively. In developing the alphabet used in a transcription scheme, therefore, one has to do one's best to refrain from making use of symbols other than those of the English alphabet.

1-2- Objectives

According to what was stated above, a comprehensive scheme for phonemic transcription of Persian in the English alphabet is expected to meet the need to transcribe Persian proper names in contexts where other alphabets are undesired or disallowed. Furthermore, this scheme should fulfill the need to transcribe Persian pieces of language in the English alphabet where the original alphabet is unavailable for non-linguistic reasons, and to transcribe Persian texts in the English alphabet as an initial teaching alphabet, in teaching Persian mainly but not exclusively as a foreign language.

1-3- Questions

While there are also quite specific questions, what follows is a list of the more general ones that might be raised in order to answer the more general question inquiring what a desirable transcription system for Persian can be:

– How should segmental elements be represented?

– How should phonological processes be represented?

– How should non-phonological phenomena be represented?

1-4- Definition of Terms

One of the key concepts in the present thesis is "alphabet". It should not, therefore, be taken for granted that it has to be distinguished from the two related terms "script" and "orthography". The technical distinctions are clarified below:

1-4-1- Alphabet versus Script

In Richards and Schmidt (2002: 21), under the entry "alphabet", you read "The English alphabet uses roman script". A script, therefore, can form the basis of an alphabet. It should be added that wherever throughout the thesis the term "Roman-based alphabet" is used, an alphabet using Roman script is being referred to.

1-4-2- Alphabet versus Orthography

In Trask (1999: 219), under the entry "orthography", you read:

> A given orthography consists of a particular version of a particular **writing system** (in the case of English, a version of the Roman alphabet employing both small letters and capital letters), a standardized system of **spelling**, often a system of word-division (in English, by white spaces), and (in almost all cases today) a particular system of **punctuation**. Often there are additional devices, such as (in English) the use of the arabic numerals (1, 2, 3, ...) for writing numbers, the specialized use of certain *symbols* (such as +, =, &, %, $, @), the presence of certain *abbreviations* (such as *Dr*, *St*, *etc.*, *e.g.*), and the use of distinctive typefaces like italic and boldface....

An alphabet, therefore, can be considered as a part of, along with all other parts of, an orthography.