Kingdom of Saudi Arabia

Ministry of Higher Education

King AbdulAziz University

College of Arts & Humanities

Dept of European Languages and Literature

أثر الاقتراض اللّغوي على الكتابة باللغة العربية

The Effect of Borrowing on the Orthography of Arabic

Leyla Abdulkadir Hussein Hashim

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Arts in Linguistics

1427 H / 2006 G

Abstract

The present study is designed to study the effect of borrowing on the orthography of Arabic. The main hypotheses to be tested were: Arabic orthography is changing to accommodate the written representation of loanwords; native graphemes are used to represent loanword phonemes; additional graphemes are borrowed from other scripts to fill the gap caused by the absence of certain graphemes in Standard Arabic; some graphemes represent more than one phoneme; diacritics are rarely used in representing loanword phonemes; native speakers of Arabic rely on the phonological (aural) route in graphemically representing loanwords which is the method whereby a word is written as it is perceived by the native speaker of the host language. The data, which consisted of 397 entries, were collected from different sources: magazines, newspapers, billboards etc…. The results indicate that native speakers of Arabic employ the aural method more than the visual method in representing the phonemes of borrowed words and that Arabic orthography is adequate and completely capable of representing the myriad of borrowed words in diverse fields from technology and medicine to food and fashion. Some graphemes that have been borrowed have not been standardized yet, for example the < ڤ > is not found on standard keyboards. Its fate remains to be seen. The effect of borrowing has not been felt in formal Arabic but is certainly present in substandard usage and in peripheral phenomena like in ads for visual effects. The purported “defectiveness” of Arabic is what has actually enabled it to encode a lot of phonological information using the minimum number of graphemes.

مستخلص

صمم هذا البحث لدراسة أثر الاقتراض اللغوي على الكتابة باللغة العربية. أما الفرضيات الأساسية التي تم فحصها هي: الإملاء العربي في طور التغيُّر؛ لاحتواء الاقتراض اللغوي. وتستخدم الحروف الهجائية للتعبير عن الأصوات المقترضة . وتستعار حروف هجائية إضافية من المجموعات الأخرى لملأ الفراغ (أو النقص) الحاصل بسبب غياب أحرف معنية في العربية. وتمثل بعض صيغ الإملاء أكثر من حرف واحد. وتستخدم الحركات الإملائية في تمثيل الحروف الهجائية المقترضة ، وذلك في النادر . ويعتمد متحدثو اللغة العربية الأصليون على المسار الصوتي للأحرف الهجائية في الإملاء ؛ لتمثيل الكلمات المقترضة . وهي الطريقة التي يتم فيها كتابة الكلمة كما تم اقتباسها من المتحدثين الأصليين للغة المضيفة. وتم جمع مادة البحث (المعلومات) من مصادر مختلفة، مثل الجرائد والمجلات والصحف واللوحات الإعلانية وخلافها. تشير النتائج إلى التالي: يستخدم الناطقون باللغة العربية المسار الصوتي أكثر من المسار المرئي لتمثيل أصوات الكلمات المقترضة. قواعد الإملاء في اللغة العربية غنية بحيث أنها قادرة تماماً على تمثيل الكلمات المقترضة على إختلافها في جميع المجالات، ابتداء من التكنولوجيا والطب وانتهاء بالغذاء والموضة. بعض الحروف الهجائية المقترضة لم يتم انتشارها بعد. فعلى سبيل المثال حرف ال<ڤ> لم يتم إدراجه كحرف هجاء أساسي في لوحة المفاتيح . إنَّ تأثير الاقتراض اللغوي غير ملموس في اللغة العربية الفصحى، ولكنه ظاهر في الاستخدامات الجانبية؛ لغرض إحداث تأثير مرئي كما هو الحال في الإعلانات. وهذا البحث يثبت مقدرة حروف اللغة العربية على تمثيل الكثير من الظواهر الصوتية باستخدام أقل عدد موجود من الحروف وهذا ينفي ما قيل عنها من نقص وقصور.

TABLE OF CONTENTS

ABSTRACT ..... **I**

مستخلص ..... **II**

ACKNOWLEDGMENTS ..... **III**

1. CHAPTER ONE **1**

1.1 BACKGROUND OF THE STUDY 1
1.2 OBJECTIVES OF THE STUDY 5
1.3 SIGNIFICANCE OF THE STUDY 6
1.4 RESEARCH QUESTIONS 8
1.5 HYPOTHESES 8
1.6 OUTLINE OF THE THESIS 9

2. CHAPTER TWO **10**

2.1 THE DIFFERENCE BETWEEN WRITING AND SPEECH 12
2.2 THE CORRELATION BETWEEN WRITING AND SPEECH 13
2.3 WHAT WRITING REPRESENTS 14
2.4 WHAT IS A WRITING SYSTEM 14
2.5 THE TYPOLOGY OF WRITING SYSTEMS 15
2.6 THE ARABIC WRITING SYSTEM 16
2.7 THE ORTHOGRAPHIC SYSTEM OF ARABIC 17
2.8 PHONETIC SYMBOLS 18
2.9 THE ARABIC SCRIPT AND THE QUR'AN 20
2.10 REFORM OF THE ARABIC SCRIPT 21
2.11 NAQT OR POINTING SYSTEM 22
2.12 I'JAM OR DISAMBIGUATION 22
2.13 SHAKL OR MARKING OF SHORT VOWELS 22
2.14 BORROWING AS A FACTOR OF CHANGE IN WRITING SYSTEMS AND LANGUAGE 23
2.15 CHARACTERISTICS OF BORROWING 24
2.16 WHY DO LANGUAGES BORROW? 25
2.17 DEGREE OF ASSIMILATION 26
2.18 ROUTES OF BORROWING 28
2.19 THE EFFECT OF BORROWING 29
2.20 LEXICON 30
2.21 MORPHOLOGY 31
2.22 SYNTAX 31
2.23 SEMANTICS 31
2.24 ORTHOGRAPHY 32

3. CHAPTER THREE **36**

3.1 DATA COLLECTION 36
3.1.1 Sources of the data *36*
3.1.2 Methodology *37*

4. CHAPTER FOUR **40**

4.1 ROUTE INTO THE LEXICON 40
4.1.1 The visual route *41*
4.1.2 The aural route *41*
4.2 THE DATA 41
4.3 SOURCES OF THE GRAPHEMES 42
4.3.1 Coined graphemes *42*
4.3.2 Native graphemes *43*
4.3.2.1 Monographs 43
4.3.2.1.1 The Consonants 44
4.3.2.1.1.1 The Stops 44
4.3.2.1.1.1.1 The voiceless bilabial stop: /p/ 44
4.3.2.1.1.1.2 The voiced bilabial stop /b/ 45
4.3.2.1.1.1.3 The voiceless alveolar stop /t/ 46
4.3.2.1.1.1.4 The voiced alveolar stop /d/ 46

4.3.2.1.1.1.5 The voiceless velar stop /k/ 47
4.3.2.1.1.1.6 The voiced velar stop /g/ 48
4.3.2.1.1.2 The Fricatives 50
4.3.2.1.1.2.1 The voiceless labiodental fricative /f/ 50
4.3.2.1.1.2.2 The voiced labiodental fricative /v/ 50
4.3.2.1.1.2.3 The voiceless alveolar fricative /s/ 51
4.3.2.1.1.2.4 The voiced alveolar fricative /z/ 52
4.3.2.1.1.3 The Affricates 53
4.3.2.1.1.3.1 The voiceless alveo-palatal affricate /č/ 53
4.3.2.1.1.3.2 The voiced alveo-palatal affricate / ǰ / 54
4.3.2.1.1.4 The Nasals 55
4.3.2.1.1.4.1 The voiced bilabial nasal /m/ 55
4.3.2.1.1.4.2 The voiced alveolar nasal /n/ 56
4.3.2.1.1.4.3 The voiced velar nasal / ŋ / 56
4.3.2.1.1.5 The Approximants 57
4.3.2.1.1.5.1 The voiced alveolar lateral approximant /l/ 57
4.3.2.1.1.5.2 The voiced alveolar central approximant /r/ 58
4.3.2.1.1.5.3 The voiced palatal approximant /j/ 59
4.3.2.1.1.5.4 The voiced labial-velar approximant /w/ 60
4.3.2.1.2 The Vowels and Diphthongs 61
4.3.2.1.2.1 The Arabic grapheme < ا > 63
4.3.2.1.2.2 The Arabic grapheme < و > 65
4.3.2.1.2.3 The Arabic grapheme <ي> 68
4.3.2.1.2.4 No grapheme 70
4.3.2.2 Digraphs 72
4.3.2.2.1 The Arabic digraph < يه / ية > 72
4.3.2.2.2 The Arabic digraph < اي > 73
4.3.2.2.3 The Arabic digraph <وي > 74
4.3.2.2.4 The Arabic digraph <او> 75
4.3.2.2.5 The Arabic digraph <يو> 76
4.3.3 Borrowed Graphemes 77
4.3.3.1 Monographs 77
4.3.3.1.1 Consonants 77
4.3.3.1.1.1 Stops 77
4.3.3.1.1.1.1 The grapheme <پ> 77
4.3.3.1.1.1.2 The grapheme <چ> 78
4.3.3.1.1.2 The Fricatives: 78
4.3.3.1.1.2.1 The grapheme <غ> 78
4.3.3.1.1.2.2 The grapheme <ڤ> 79
4.4 DEPTH OF ORTHOGRAPHY 79
4.5 DIACRITICS 81
4.6 MORPHOSYNTACTIC LEVEL 82
4.6.1 Acronymy 82
4.6.2 Polyvalence: bivalence 84
4.6.3 Graphological mirroring 84

5. CHAPTER FIVE 85

5.1 FINDINGS 85
5.1.1 Route into the lexicon 85
5.1.2 Sources of the graphemes 85
5.2 RECOMMENDATIONS FOR FURTHER RESEARCH 87

REFERENCES 88

APPENDIX A 91

ملخص 99

1. CHAPTER ONE
INTRODUCTION

1.1 Background of the Study

Writing is estimated to be around 5000 years old in comparison to speech which is as old as Man himself (Coulmas 1996: 481, Daniels 33, Sampson 47). A fundamental difference between speech and writing is that speech is acquired by all those who are exposed to language at an early age and are not hindered by, for example, deafness. Writing on the other hand is an artificial skill that needs to be learned. In addition, spoken language changes and evolves "without the conscious interference of its speakers but writing can be petrified, or reformed, or adapted or adopted at will" (Daniels 2). Historically writing was learned by an elite minority who had access to literacy mainly through religion. As Daniels puts it, "All humans speak; only humans in civilizations write, so speech is primary and writing is secondary" (1). Even today literacy is a measure of civilization with governments and the UNESCO trying to eradicate illiteracy. But writing as a field worthy of scientific study has been ignored early in the 20th century as a reaction to the former emphasis placed on writing to the exclusion of speech. Linguists were busy prescribing correct forms to be used that mirrored the rules and standards of writing because those who possessed the skill to write were in a minority. Therefore, speaking like a book was accorded prestige. When linguists encountered exotic unwritten languages spoken by primitive peoples the tide changed and emphasis was placed on studying the oral production of those indigenous tribes. Writing had no place in linguistics because within the framework of modern linguistics the

corpus of data from which native speakers abstract their hypotheses is spoken language (Sampson 13). Ironically, those who professed this view were oblivious to the fact that the object of their study was the language in transcribed form. They suggest theories about "spoken language *written down*" (Daniels 11). The tide turned again and it became clear that writing has an effect on how we perceive and how we process language. While writing had been studied earlier in the last century for its obvious practical implications for the teaching of reading and writing, spelling reforms, and devising orthographic systems; after the introduction of computational linguistics, it is also being studied for its implications for Natural Language Processing, Text-to-Speech and Speech-to-Text conversion, synthesized speech, information systems and data retrieval systems, to name just a few. Having said this, the question that remains is: What is writing? I shall attempt to answer this question by stating first what writing is not. Writing is not transcription. It is not a faithful rendering of the sounds of language into another medium—this time a graphic one. Although the International Phonetic Alphabet (IPA) is a modern writing system devised to relate graphemes to phonemes in an unambiguous manner, it was never meant for everyday use. "The IPA presupposes that sounds can be counted and tries to approximate a bi-unique mapping relation of 'one sound, one letter', although the Principles of the International Phonetic Association warns that 'vowel symbols are necessarily elastic in their values'. It defines phonemes as families of related sounds and admits that 'it is not possible to design letters for the representation of all distinguishable shades of sound' " (qtd. in Coulmas 2003: 30). Because the IPA has been

graphically based on the Latin alphabet it “has also the incalculable disadvantage of fostering a conceptual confusion of writing and transcription [which resulted] in [the] faulty and misleading conjecture . . . that the Latin alphabet and its Greek precursors were transcription systems and that whatever discrepancies between sounds and letters found in Greek and Latin texts are simply a result of sound change over time“ (Coulmas 2003: 30). Now, to answer the question: What is writing? Writing can be defined as a form of graphic communication where information is related without providing the reader with minute phonetic details because the reader knows the language that is written and does not depend on such details for deciphering the meaning encoded in the text.

Although modern linguistics denied the value of writing and orthography as systems fit for scientific inquiry because of the primacy of speech mentioned above, one cannot help but notice that orthography does influence speech and vice versa. Evidence for this lies in what is termed in the literature as “spelling pronunciation” and “pronunciation spelling”. "Spelling pronunciation" (also called the visual route) is a process whereby the graphemes used to write a loanword are based on the actual spelling of the loanword in the lending language whereas "pronunciation spelling" (also called the aural route) is a process whereby the graphemes chosen to represent the loanwords are based on the phonetic shape of the word in the lending language or the phonetic shape of the word in the host language based on the degree of assimilation of the word.

All living natural languages are dynamic in that they change and evolve. One of the factors that bring about language change is language contact and the resultant borrowing. In turn scripts evolve because they come to be used for a language for which they were not invented. "The evolutionary development of scripts including the alphabet is the simple consequence of attempting to use a graphic system invented to be 'read' in one language for which it is thereby reasonably suited, to convey messages to be 'read' in another language for which it is not well suited" (Olson 67). Arabic is one such script that is evolving for it is being used to represent sounds for which it was not originally devised. The purpose of this study is to track these changes.

The writing system of Arabic is classified as an abjad. An abjad is a writing system that denotes only consonants (Daniels 4). But one of its distinguishing features is that although all its graphemes represent only consonants, there are three graphemes that double as long vowels. These are the <أ ، ي ، و> which were originally the graphemes for the glottal stop /ʔ/, the alveopalatal glide /j/ and the bilabial glide /w/, respectively. Short vowels are typically not written even though much morphological and grammatical meaning is encoded in them. Only the roots and stems of an inflected word are written, and the burden of deciphering the written code is left to the reader who relies on context for that. Short vowels, diacritics and gemination marks are represented in Standard Arabic in religious texts, poetry, elementary textbooks and children's literature or when there is a need to disambiguate a word.

As a result of increasing globalization and language contact, the Arabic lexicon is colored with loanwords in every field from science and technology to food and fashion. For example, computer كمبيوتر , fax فاكس , telephone تلفون , modem مودم , ice tea آيس تي , cappuccino كابتشينو , espresso اسبريسو , pizza بيتزا , croissant كروسان , mall مول , supermarket سوبر ماركت, bank بنك , crepe كريب , pullover بلوفر , T-shirt تي شرت , and séchoirسشوار , to name just a few. Some have already been accepted by the Arabic Language Academy as Arabicizations such as " كمبيوتر، مودم، تكنلوجيا"; others have not. Nevertheless, whether reading a book, newspaper, magazine, billboard, or business card one cannot help but notice the peculiar orthography of some non-native words.

Loanwords pose a challenge to the writing system since they create a gap between the existing graphemic system of the host language and the phonemes in the loanwords. Many of the loanwords that are being imported into Arabic today have phonemes that are not part of the phonemic inventory of Standard Arabic.

In this study I have attempted to examine the effect of loanwords on the orthography of Arabic. In particular, I have described the strategies that native speakers developed to represent loanwords graphemically and, as a result, have modified the structure of Arabic orthography.

1.2 Objectives of the Study

The main objective of the study is to analyze the strategies the native speakers of Arabic would normally employ in representing the non-native phonemes in the loanwords.

To achieve this, the study investigates the source of the graphemes used in representing loanwords. It pinpoints whether:

1. the graphemes are coined (new graphemes are invented),
2. the graphemes are borrowed (from other scripts from the same Afro-Asiatic family that Arabic belongs to like Persian),
3. the present stock of graphemes is reanalyzed (where new values are assigned to existing graphemes where one grapheme represents more than one sound), or
4. rearranged (where new digraphs or trigraphs are used to represent the sounds in the loanwords).

The study has, in addition, identified the route the words are taking into the lexicon. The two routes of interest to this study are:

1- the visual route (also called spelling pronunciation) whereby the graphemes used to write the word are based on the actual spelling in the lending language, and

2- the aural route (also called pronunciation spelling) whereby the graphemes chosen to represent the loanwords are based on the phonetic shape of the word in the lending language or the phonetic shape of the word in the host language based on the degree of assimilation of the word.

1.3 Significance of the Study

The study provides insights into linguistic change in progress.

It sheds light on how native speakers innovatively manipulate the present script to represent sounds that have no corresponding graphemes.

The study is used to deflect the numerous “accusations that have been directed against Arabic orthography in the name of cultural development” cited in a paper entitled “A Linguistic Study of the Adequacy of Arabic Orthography”, where Abderrahman (1995) quotes:

1- Fahmi (1944) who thought that “only the phonetic writing of the Latin could save the Arabic script from ruin”.
2- Mousa (1955) who saw that Fahmi’s “proposal for Latinizing the Arabic script is a panacea for our moral, cultural and spiritual values” and without which “we (the Arabs) would [not be able] to welcome modern industrial civilization, with its moral, cultural and spiritual values”.
3- Taha Husain (1960) who shared the view that “the Arabic writing system is so deficient that it should be replaced by the Latin script”.
4- Ben Salama (1971) who believes that the “use of a defective script has prevented the Arabic language from entering the age of the printing press; that as long as it is denied this privilege, there will be no real literary production and Arabic will remain the difficult language many believe it to be”.
5- Mahmoud (1979) who claims that the “feeling among educators had become so strong that the prodigious efforts they were exerting in filling the terminological gaps, simplifying the structure of the language, and Arabicizing the subjects were being constantly thwarted by a defective script”.

These claims are rebutted by showing that the Arabic script is neither inadequate nor defective and can be used to put into writing the flood of loanwords from medicine and technology, to food and fashion.

1.4 Research Questions

In light of the above, the study answers the following questions:

1. How are the phonemes (consonants, long vowels and short vowels, and diphthongs) in the loanwords represented?
2. Are native graphemes used to represent the phonemes of the loanwords?
3. Are different graphemes borrowed from other scripts to fill the gap caused by the absence of certain graphemes in Arabic?
4. Are diacritics used? If so, when? how?
5. Is the structure of Arabic orthography changing as a consequence of these strategies?

1.5 Hypotheses

The following hypotheses are proposed in relation to the above research questions:

1. Arabic orthography is changing to accommodate the written representation of loanwords.
2. Native graphemes are used to represent loanword phonemes.
3. Additional graphemes are borrowed from other scripts to fill the gap caused by the absence of certain graphemes in Standard Arabic.
4. Some graphemes represent more than one phoneme.

5. Diacritics are rarely used in representing loanword phonemes.
6. Native speakers of Arabic rely on the phonological (aural) route in graphemically representing loanwords which is the method whereby a word is written as it is perceived by the native speaker of the host language.

1.6 Outline of the Thesis

Chapter I is an introduction. The theoretical framework of the thesis is introduced in Chapter II. Chapter III includes a discussion of the data sources, and the collection and organization process. Chapter IV deals with the data analysis and findings. Finally, Chapter V presents the conclusion and recommendations for future research.