David R. Lawson. An Evaluation of Arabic Transliteration Methods. A Master's Paper for the M.S. in L.S degree. April, 2008. 55 pages. Advisor: Ronald Bergquist

The American Library Association and the Library of Congress currently use a cooperatively developed Arabic transliteration system that is not ASCII-compatible and that incorporates the use of diacritical marks native to neither Arabic nor English. This study seeks to investigate whether the adoption of an alternate Arabic transliteration system by ALA and LC can increase both user access to the materials as well as the ability of librarians to correctly catalog them. The various systems are evaluated based upon phonetic and spelling accuracy, as well as usability, the adherence to not using non-native diacritics, and their compatibility with ASCII standards. A parallel with the issues in Korean transliteration is made in order to show how another language written in a non-Roman script approached the issues. After the analysis, a recommendation is made and avenues for further study are explored.

Headings:

Arabic language/Transliteration

Arabic literature/Cataloging

Cataloging

Cataloging/Transliteration/Use Studies

Transliteration

AN EVALUATION OF ARABIC TRANSLITERATION METHODS

by
David R. Lawson

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

March 2008

Approved by

_____

Ronald Bergquist

TABLE OF CONTENTS

INTRODUCTION

Transliterating a foreign script into Roman characters is a complex process that requires adherence to the original language's phonetics within the confines of the English alphabet. Many non-Romance languages include numerous sounds not native to English, nor easily represented within the standard 26 characters used in English. This makes it difficult to accurately express in Roman script the sounds and letters of a language that originates from somewhere other than Western Europe.

Because a number of patrons likely to use an Anglophone library's OPAC (Online Public Access Catalog) have been exposed to an Indo-European language other than English, it is rare for a cataloger to have to take steps to ensure that someone can understand an acute diacritic (́) in French or a tilde (~) in Spanish. However, knowing what a bracket sitting at a ninety degree angle above a letter means, ǧ, or what a dot below a letter, ṣ, is meant to convey, can be a very different story.

Adding to the problem is the idea that creators of a transliteration system may have conflicting ideas of what the objective of such a system should be. For linguists, diacritical marks that appear foreign to most people would be an easy read for them, and would be necessary to indicate exactly how a letter is to be correctly pronounced. For a computer programmer, a system that adheres to the American Standard Code for

Information Interchange (ASCII) would suffice. For a cataloger, something that a non-expert on a language can create, without using characters that will not display in an OPAC, would be a good start. And for library patrons, something that they could read without a primer for diacritical marks would be quite useful.

With this many competing perspectives in the world of transliteration, it is of little surprise that more than a dozen systems exist in the Arabic language alone. A few of the more prominent ones, listed in a random order, are:

- International Standard (ISO) 233-2
- Deutsche Morgenländische Gesellschaft (DMG)
- Standard Arabic Technical Transliteration System (SATTS)
- Deutsches Institut für Normung (DIN) 31635
- Buckwalter
- Royal Jordanian Geographic Centre
- British Standards Institute (BS) 4280
- United Nations Experts on Groups of Geographical Names (UNGEGN)
- Qalam
- American Library Association-Library of Congress (ALA-LC)
- Spanish Arabists School (SAS)
- Arabizi
- Survey of Egypt System

All of the perspectives mentioned before the list are to some degree represented in at least one of these transliteration systems. The unique aspects of Arabic are in many ways the cause of this proliferation.

Arabic characters can have up to four different forms, with the shape of the character dependent upon the letter's placement within the word. The placement types are initial, medial, final, and isolated. Many of these letters do not translate well to an untrained English speaker, as some of these letters, and the sounds they produce, have no Roman characters that act as accurate phonetic matches. This lack of easy transition from Arabic consonant sounds to English consonant sounds has paved the way for such discrepancies between transliteration systems to emerge.

Vowels play a less important role in the divisiveness, but they deserve mention in order to better understand what part, or lack thereof, they play in a Romanization scheme.

There are five long vowels in Arabic and they are all printed in standard writing along with the consonants. However, the three short vowels, along with the "sukun", which indicates the absence of a vowel sound between consonants, are not printed. It is assumed that a person literate in the language will know what short vowels are to appear.

An English example would be bsktbll. The short vowels are not printed, but it is obvious to a native speaker that the word is "basketball". There are some notable exceptions to this "no short vowel" rule in Arabic – book covers, title pages, primers for young students, and holy texts amongst the most important – but catalogers and patrons will rarely find Arabic monographs or serials with full diacritics throughout.

The need for this evaluative study becomes apparent when looking at the substantial list of transliteration options, with no one system particularly dominant in terms of overall prevalence. The reason all these systems still exist is because they are all still in use. Thus, some sort of process to determine the relative strengths and weaknesses of them must be employed not only for the good of library patrons and catalogers, but also for participants in other fields where non-Roman transliteration is in any way related to the profession.

Arabic is especially important at this moment in American, and world, history. Since 2001 the demand for Arabic speakers and Arabic literature has increased exponentially. However, this study seeks to evaluate methods that would primarily aid Arabic novices and those who have no ability to read this language. So how do transliteration methods and the need for Arabic ability relate?

The most important relation would be access to materials. Currently, because OCLC (Online Computer Library Center) uses the ALA-LC/UNGEGN transliteration method, local ILS (Integrated Library System) platforms such as Millennium, Aleph, and Voyager use that method as well. This means that if patrons conduct an author or title search using Roman characters, the results are almost guaranteed to be incomplete or blank altogether, as names written in ALA-LC/UNGEGN will not necessarily be found through a query that uses a non-diacritics based transliteration. This is critically important to library patrons conducting any sort of search through an OPAC. It literally means that typing an Arabic writer's name in, verbatim in Roman characters, may not result in a complete list of his available works being displayed.

Thus, this paper will examine the strengths and weaknesses of half a dozen Arabic transliteration systems. They will be evaluated for accuracy and usability. Ultimately, this could prove useful to library patrons and universities by providing them with an objective, ranked look at which transliteration systems do the best job of adhering to correct Arabic pronunciation while using characters that will properly display on the screen.

LITERATURE REVIEW

The most important aspect of the literature review portion of the research for this paper does not focus on similar studies involving the merits of Arabic transliteration systems, as there are none in existence, but instead upon how the situation that now exists came to be. Ultimately, with an adequate, but hardly overwhelming quantity of articles devoted to subjects tangential to Arabic transliteration, mostly covering the difficulties of cataloging in the language, the literature review is more of a historiography that seeks to identify the rhyme and reason behind the current state of affairs. Finding the rationale for the decisions that led the American Library Association to create an Arabic transliteration system that a substantial portion of its member libraries cannot display on an OPAC, and that university catalogers cannot effectively use without constantly turning to character insertions in their ILS services, became a priority. Why did ALA-LC choose to go one way while the computer world and corporate America went another? And in the midst of all this, which way did native Arabic speakers choose to proceed?

It is interesting to note that the current system did not have its beginnings at an ALA conference, a Library of Congress memo, or even a meeting of Arabic catalogers. Instead, the decree that has landed the Arabic transliteration world in its current predicament came from the United Nations.

In 1972 the United Nations Experts on Groups of Geographical Names approved the system created by a gathering of Arabic experts in Beirut a year earlier (UNGEGN Working Group, 2003). And though the author could not find information directly linking the ALA-LC method to the UNGEGN method, the fact that the two systems' character transliterations are completely identical, save for the shape of the diacritic beneath the ص, ض, ط, ظ, ح, is very compelling evidence that the similarities are more than coincidental.

However, at the time the UNGEGN transliteration method was created, OCLC did not support the Arabic script on its WorldCat (OCLC's union catalog) interface. This did not happen until 1992, when the Research Libraries Group finally did so, years after Hebrew and Cyrillic and nearly a decade after Chinese, Japanese, and Korean (Eilts, 1995).

Of course, the UN-ALA-LC-OCLC (termed ALA-LC within the text henceforth) work did not take place in a bubble. Others were working on their own methods with their own agendas that did not necessarily follow the linguist-derived code already set forth.

After all, "in any discussion of Arabic Romanization, one is entering a field where there is little practical or terminological agreement" (Beesley, 1998). Many of the disagreements focus on morphological analysis and other linguistic and natural language phenomena outside this author's area of competence. However, the aspects of the transliteration problems that do not require expertise in the field of computational linguistics are easy to recognize. The ALA-LC model posed an even larger problem

fifteen years ago than it does now. Far fewer OPACs and ILS interfaces could display the diacritics in a Romanized script, much less the Arabic vernacular. Thus, from the 1972 launch through ALA's 1997 minor revision, there was plenty of room for something more practical to come along (Barry, 1997).

METHODOLOGY

Each system will be evaluated by two broadly defined components crucial to the success of any transliteration method for non-Roman scripts – accuracy and usability. To provide an example of how the Arabic systems will be graded, a brief examination of Korean transliteration methods is presented here.

The Korean Transliteration Experience

Arabic is of course not the only non-Roman language that has seen several transliteration methods attempt to make it legible for an Anglophone populace. For Korean, there are four transliteration systems worth noting:

1. The 1937 McCune-Reischauer system, created by a pair of graduate students. The modern ALA-LC method is almost a direct copy.
2. The 1942 Yale system, no longer in use outside linguistic circles
3. The 1992 North Korean method
4. The 2000 Revised Romanization scheme from South Korea (UNGEGN, 2003)

Of these four, two are particularly relevant to the following discussion of Arabic transliteration, the McCune-Reischauer (MCR) system

McCune-Reischauer

Hangul	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
ㅏ	a	✓	✓	✓
ㅑ	ya	✓	✓	✓
ㅓ	ŏ	✓	✓	✓
ㅕ	yŏ	✓	X	X
ㅗ	o	✓	✓	✓
ㅛ	yo	✓	✓	✓
ㅜ	u	✓	✓	✓
ㅠ	yu	✓	✓	✓
ㅡ	ŭ	✓	X	X
ㅣ	i	✓	✓	✓
ㅘ	wa	✓	✓	✓
ㅝ	wŏ	✓	X	X
ㅐ	ae	✓	✓	✓
ㅔ	e	✓	✓	✓
ㅚ	oe	✓	✓	✓
ㅟ	wi	✓	✓	✓
ㅢ	ŭi	✓	X	X
ㅙ	wae	✓	✓	✓
ㅞ	we	✓	✓	✓
ㅒ	yae	✓	✓	✓
ㅖ	ye	✓	✓	✓
ㅂ	p	X	✓	✓
ㅈ	ch	X	✓	✓
ㄷ	t	X	✓	✓
ㄱ	k	X	✓	✓
ㅅ	s,sh	X	✓	✓
ㅁ	m	✓	✓	✓
ㄴ	n	✓	✓	✓

McCune-Reischauer

Hangul	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
ㅋ	k'	✓	✓	✓
ㅌ	t'	✓	✓	✓
ㅊ	ch'	✓	✓	✓
ㅍ	p'	✓	✓	✓
ㄹ	l,r	✓	✓	✓

and the Revised Romanization (RR) scheme

Revised Romanization

Hangul	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
ㅏ	a	✓	✓	✓
ㅑ	ya	✓	✓	✓
ㅓ	eo	✓	✓	✓
ㅕ	yeo	✓	✓	✓
ㅗ	o	✓	✓	✓
ㅛ	yo	✓	✓	✓
ㅜ	u	✓	✓	✓
ㅠ	yu	✓	✓	✓
ㅡ	eu	✓	✓	✓
ㅣ	i	✓	✓	✓
ㅘ	wa	✓	✓	✓
ㅝ	wo	✓	✓	✓
ㅐ	ae	✓	✓	✓
ㅔ	e	✓	✓	✓
ㅚ	oe	✓	✓	✓
ㅟ	wi	✓	✓	✓
ㅢ	ui	✓	✓	✓
ㅙ	wae	✓	✓	✓
ㅞ	we	✓	✓	✓

Revised Romanization

Hangul	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
ㅒ	yae	✓	✓	✓
ㅖ	ye	✓	✓	✓
ㅂ	b	✓	✓	✓
ㅈ	j	✓	✓	✓
ㄷ	d	✓	✓	✓
ㄱ	g	✓	✓	✓
ㅅ	s	✓	✓	✓
ㅁ	m	✓	✓	✓
ㄴ	n	✓	✓	✓
ㅋ	k	✓	✓	✓
ㅌ	t	✓	✓	✓
ㅊ	ch	✓	✓	✓
ㅍ	p	✓	✓	✓
ㄹ	l,r	✓	✓	✓

(Romanization, 2000)

Much like Arabic, a balance has to be found between phonetic accuracy and computer-based usability. But unlike Arabic, there is no longer any real debate over the superior system. As the shaded cells in the above tables illustrate, McCune-Reischauer (MCR) uses diacritics not native to English or Korean. Revised Romanization (RR) does not. MCR has several characters that are not ASCII compatible. RR has zero. And every single one of MCR's diacritical marks maintains its phonetic accuracy when replaced by a double letter combination.

Despite this evidence, and despite the fact that the South Korean government has used nothing but the RR system since its inception in 2000, MCR is still in use. Its most

notable practitioners include the American Library Association, the Library of Congress, the United Nations (though they have at least discussed upgrading), and the Online Computer Library Center.

Accuracy

Accuracy, for the purposes of this paper, will be an examination of the method's adherence to the pronunciation of the original Arabic letter. In other words, how close does the Roman letter(s) or combination of letter and diacritical mark come to representing the original pronunciation of the Arabic character? Systems that use diacritics will be evaluated for phonetic and spelling accuracy without penalty.

Things can of course get tricky in this portion of the evaluation process. A good example would be the letters ض and ص written with the Q and W, respectively, on a keyboard. The first letter listed sounds not all that different from د, which every transliteration system agrees is a one-to-one match for the Roman letter D. However, ض has no direct English equivalent. One can make the sound by saying *daad,* which involves elongating the mouth, producing a stressed D sound, then a AH, equivalent to the vowel sound in "ball", and then finally popping the tongue off the roof of the mouth for an aspirated D. It is rather difficult to convey those instructions through the use of a single key.

Thus, with no direct English equivalent, Arabic-to-English transliteration systems have taken different approaches to this letter. One of them, Qalam, has used a capital D to emphasize the stressed sound. Another, SATTS, in keeping with its policies of making

adherence to ASCII compatibility and not repeating characters more important that phonetic matching, uses a V. All the rest use a lowercase d with some sort of diacritical mark under the letter. These marks are native to neither English nor Arabic; they are not ASCII compatible; and they often fail to display properly in an OPAC.

Arabic Letter	Name of Letter	ISO 233-2	Qalam	SATTS	Arabesh	Buckwalter	ALA /LC
ض	daad	ḍ	T	V	D/9'	D	ḍ

Another good example of the problems transliteration systems face is the character غ, produced by typing the letter Y on the keyboard. The isolated, initial, medial, and final forms of this letter look very different from one another, as displayed here, in left-to-right order, غ غس مغى مغ with the characters being attached to other letters in the three non-isolated cases.

Correctly pronouncing this letter involves closing the throat, placing the tongue far back against the roof of the mouth, and then attempting to say a guttural *gh'ayn,* while forcing the air out at the same time the tongue is dropped back to a normal position. The difficulties of transferring the instructions of the previous example to a standard keyboard look fairly tame compared to this one.

So, what is a transliteration system to do with this matter? SATTS uses a capital G; Qalam and ALA-LC use a gh; Buckwalter uses a g; and ISO 233-2 uses a g with a diacritical mark above it.

Arabic Letter	Name of Letter	ISO 233-2	Qalam	SATTS	Arabesh	Buckwalter	ALA
غ	gayn	ḡ	gh	G	gh/3'	g	gh

As one can see, there is some disagreement, but a general consensus on the sound to be made, which unfortunately comes nowhere close to being accurate. Thus, some allowances have to be made for any Arabic transliteration system, as on several occasions no combination of Roman characters or marks can accurately convey the correct pronunciation of a letter.

Further adding to the difficulty is the fact that Moroccan Arabic is as close to Iraqi Arabic as English is to Swedish. The consonant roots and the verb formations key most educated people into seeing there is some common basis, but the dialects are so far apart that transliterated names are often written with completely different Roman characters. The g used for the *ghayn* character above may actually cause problems for Egyptians, as they pronounce ج with a hard 'g' sound ("Geem" with a G as in "get"), whereas all other Arabic-speaking countries use a 'j' sound ("Jeem" with a J as in "jet"). Thus, it is easy to see that not only do the Arabic sounds not native to English make a perfect transliteration system a near impossibility, but the dialect variations within Arabic itself may very easily cause a system's strength to vary, to at least some degree, from country to country.

Examples of this problem can be seen in NACO (Name Authority Cooperative Program) authority files for past Arabic writers, some of whom have seen their names transliterated any number of ways. These problems of course do not begin and end with the Library of Congress, as other government agencies face difficulties concerning this same matter. Mohammed bin Sayed may be wanted in three countries, but Mohammad bin Saeed might be a respected businessman from Tunis. This problem reaches a level of

absurdity with the name of the acting head of Libya, as there are at least 32 different ways to transliterate his name (Variations, 2004):

1. Muammar Qaddafi
2. Mo'ammar Gadhafi
3. Muammar Kaddafi
4. Muammar Qadhafi
5. Moammar El Kadhafi
6. Muammar Gadafi
7. Mu'ammar al-Qadafi
8. Moamer El Kazzafi
9. Moamar al-Gaddafi
10. Mu'ammar Al Qathafi
11. Muammar Al Qathafi
12. Mo'ammar el-Gadhafi
13. Moamar El Kadhafi
14. Muammar al-Qadhafi
15. Mu'ammar al-Qadhdhafi
16. Mu'ammar Qadafi
17. Moamar Gaddafi
18. Mu'ammar Qadhdhafi
19. Muammar Khaddafi
20. Muammar al-Khaddafi
21. Mu'amar al-Kadafi
22. Muammar Ghaddafy
23. Muammar Ghadafi
24. Muammar Ghaddafi
25. Muamar Kaddafi
26. Muammar Quathafi
27. Mohammer Q'udafi
28. Muammar Gheddafi
29. Muamar Al-Kaddafi
30. Moammar Khadafy
31. Moammar Qudhafi
32. Mu'ammar al-Qaddafi

It could be worse. Not a single name on that list uses non-ASCII diacritics. Had those been included as well, the tally would have topped fifty. Thus, ultimately, these points illustrate the need for finding a universally adopted transliteration method, one which will inevitably have to balance accuracy with usability.

Usability

The usability portion of the evaluation process will examine each system's use or non-use of diacritical marks non-native to either language in question, as well as each system's adherence to ASCII standards and, in turn, compatibility to OPACs.

When patrons or catalogers encounter a letter with an unfamiliar diacritical mark above, beneath, or beside it, it is logical to assume they will probably ignore it. The second most likely outcome is that they will recognize the mark as a modifier, but will not know how it changes the pronunciation of the letter in question. This makes the diacritical marks not only unbeneficial, but in fact potentially harmful.

Why is this harmful? It is harmful because non-ASCII diacritics do not display correctly on many OPACs used by libraries. Take, for example, these two bibliographic records, the first of which is broken across two pages:

Uniform title	Kalīlah wa-Dimnah
Title	**Kitāb Kalīlah wa-Dimnah / tālīf Bīdbā al-Fīlisūf al-Hindī ; tarjimah ilá al-□Arabīyah fī ṣadar al-dawlah al-□abbasīyah □Abd Allāh ibn al-Muqaffa□ ; qararat Wizārat al-Ma□ārif al-□Amūmīyah bi-tārīkh 4 min rabī□ al-awwal sanat 1320 (10 min Yūnīyah 1902 raqm 896).** **كتاب كليلة و دمنة / تأليف بيدبا الفيلسوف الهندي ؛ ترجمة إلى العربية في صدر الدولة العباسية عبد الله بن المقفع ؛ قررت وزارة المعارف العمومية بتاريخ ٤ من ربيع الاول سنة ١٣٢٠ (١٠ من يونية ١٩٠٢ رقم ٨٩٦)**
Imprint	al-Qāhirah : Maṭb□a al-Amīrīyah bi-Būlāq, 1937. القاهرة ، مطبعة الاميرية ببولاق, ١٩٣٧.
Description	321 p. ; 24 cm.
Note	Tab□ hādha al-kitāb □alá nafaqatiha wa-tadrīsiha bi-al-madāris al-amīrīyah. طبع هذا الكتاب على نفقتها و تدريسها بالمدارس الاميرية.
WorldCat no.	123010399

Description	321 p. ; 24 cm.
Note	Tab□ hādha al-kitāb □alá nafaqatiha wa-tadrīsiha bi-al-madāris al-amīrīyah. طبع هذا الكتاب على نفقتها و تدريسها بالمدارس الاميرية.
Other author	Ibn al-Muqaffa□, d. ca. 760. ابن المقفع, d. ca. 760.
Other corporate author	Wizārat al-Ma□ārif al-□Amūmīyah. وزارة المعارف العمومية.

and

Author	□Abbās, al-□Abbās □Alī Yaḥyá. عباس, العباس علي يحيى.
Title	**Matāhāt qā□ al-madīnah : riwāyah / al-□Abbās □Alī Yaḥyá al-□Abbās.** **متاهات قاع المدينة ، رواية / العباس علي يحيى العباس.**
Edition	al-Ṭab□ah 1. الطبعة 1.
Imprint	[al-Kharṭūm?] : al-□A.□A.Y. al-□Abbās, 2006. [الخرطوم؟] : الع.ع.ي. العباس,&#x٢٠٠ئ؛ 2006.

Both of these records display correctly within the ILS and within OCLC WorldCat. However, the most up-to-date web browsers will not necessarily support all the diacritics. In these cases the ط and the ع are not displaying because their ALA-LC/UNGEGN transliterations are not Internet Explorer 5.0 compatible. In order to effectively evaluate systems all the way around, both ASCII compatibility and diacritical usage will be considered.

The research revealed no precedent for a numerical evaluation of non-Roman transliteration systems. Therefore, a new method was developed by the author.

EVALUATION

Transliteration systems under study here were selected for evaluation based upon the presence of unique attributes. Methods which incorporate non-native diacritical marks and methods that may not accurately convey the correct pronunciation of every letter were not discriminated against. However, where there are multiple methods that are identical to another one in existence, only one is listed here. The more prominent and more widely used system is named in the case of systems that are identical save for the title.

Each system receives a score from 0 to 120, with 60 points allotted to phonetic accuracy and 60 points allotted to usability. Of the latter 60, 30 points come from ASCII compatibility and 30 points come from maintaining non-use of diacritics native to neither English nor Arabic. Each point is awarded on a simple yes/no basis, i.e.

- Usability: Is this transliterated character an accurate representation of the pronunciation of the vernacular?
- Accuracy
    - Is this character ASCII compatible?
    - Does this character use non-native diacritics?

The evaluation results are to be captured in a table.

Transliteration system name

Evaluation of:	Criteria	Possible points	Points scored
Accuracy	Is this transliterated character an accurate representation of the pronunciation of the vernacular?	60	
Usability	Is this character ASCII compatible?	30	
	Does this character use non-native diacritics?	30	
Total points		120	

All transliteration systems are given credit for finding a "best fit" for Arabic characters that simply cannot be transliterated without writing out the word, such as the story of ض , ص , and غ discussed earlier. It is perhaps best to think of a non-English speaker learning the letter W. Every other English letter is a simple monosyllabic utterance that does not require multiple vowel and consonant sounds to be made, i.e. A, G, N, Y, etc. A single character is able to denote a single sound from each. In Arabic, there are just more of the W-like letters, for which a single character has to serve as the representation of more than one sound, the same as W representing "Double-you".

Thus, taking everything under consideration, the Arabic transliteration systems to be evaluated are as follows, with the first five systems (ISO 233-2, Qalam, SATTS, Arabic chat alphabet [Arabesh], Buckwalter) ordered randomly, and ALA-LC, as it is the current standard for Anglophone Libraries, coming last.

ISO 233-2

ISO 233-2 is the 1993 revised Arabic transliteration system created by the International Organization for Standards (ISO 233, 2005). When creating a transliteration system the non-governmental organization claims, "views of all interests are taken into account: manufacturers, vendors and users, consumer groups, testing laboratories, governments, engineering professions and research organizations" (ISO – Standards, 2008). The company's profile leads one to believe entities relying upon Unicode may find this method useful. However, Arab governments and U.S. government entities do not currently use this system.

ISO 233-2

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII	Non-use of Diacritics
ا	**alif**	ā	✓	✓	X
ب	**ba**	b	✓	✓	✓
ت	**ta**	t	✓	✓	✓
ث	**tha**	t̲	✓	X	X
ج	**gim**	ǧ	X	X	X
ح	**ha**	ḥ	✓	X	X
خ	**kha**	h̲	✓	X	X
د	**da**	d	✓	✓	✓
ذ	**dal**	d̲	✓	X	X
ر	**ra**	r	✓	✓	✓
ز	**zay**	z	✓	✓	✓
س	**sin**	s	✓	✓	✓
ش	**shin**	š	✓	X	X
ص	**saad**	ṣ	✓	X	X
ض	**daad**	ḍ	✓	X	X
ط	**ta**	ṭ	✓	X	X
ظ	**za**	ẓ	✓	X	X
ع	**ayn**	`	✓	✓	X

ISO 233-2

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
غ	**gayn**	ḡ	✓	X	X
ف	**fa**	f	✓	✓	✓
ق	**qa**	q	✓	✓	✓
ك	**ka**	k	✓	✓	✓
ل	**la**	l	✓	✓	✓
م	**mim**	m	✓	✓	✓
ن	**nun**	n	✓	✓	✓
ه	**ha**	h	✓	✓	✓
و	**wa**	w	✓	✓	✓
ى	**ya**	y	✓	✓	✓
ة	**ta marbuta**	h,t	✓	✓	✓
ء	**hamza**	‘	✓	✓	✓

Every non-standard character is Unicode compatible, however author tests concluded that an OPAC may fail to display up to 13 of the 30 transliterated letters above. Not surprisingly, ASCII’s 255-character range comes nowhere close to incorporating all of this, as only the ا and the ع of those 13 non-displaying characters can be properly coded. Thus, the only method left to convey the other 11 characters is to type in the code for the letter and the diacritic back-to-back. This causes a side-by-side display, i.e. ˇS instead of Š for the ش.

Despite the system’s lack of usability, it does prove to have significant value in its adherence to correct phonetics. This transliteration method lines up 16 characters one-to-one, with no diacritics. However, there should be 17. In what appears to be an effort to conform to the speech patterns of Egyptian Arabic, which is spread through the country’s

media, television, and film dominance throughout the Arab world, ISO 233-2 has used a ǧ instead of a simple j for ج

For the other 14, ISO 233-2 does a stellar job in terms of finding a way to differentiate the sound of one character from another. It separates ز from ظ as well as س from ش from ص. However, diacritics native to neither English nor Arabic are used in every case.

Thus, it becomes clear that the system the International Organization for Standards has made contains more bad in it than good. The numbers bear that out:

ISO 233-2

Evaluation of:	Criteria	Possible points	Points scored
Accuracy	Is this transliterated character an accurate representation of the pronunciation of the vernacular?	60	58
Usability	Is this character ASCII compatible?	30	19
	Does this character use non-native diacritics?	30	17
Total points		120	94

Qalam

Qalam is a morphological Arabic-Latin-Arabic transliteration system that seeks to "transliterate Arabic script for computer communication by those literate in the language" (Heddaya, 1985). It can be transliterated in both directions, by humans and by automation. It is also one hundred percent ASCII compatible and it does not incorporate a single non-native diacritic. The only non-lettered transliterations are for the ء and the ع, both of which fall into the category of non-Romanizable characters.

Qalam

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII	Non-use of Diacritics
ا	**alif**	aa	✓	✓	✓
ب	**ba**	b	✓	✓	✓
ت	**Ta**	t	✓	✓	✓
ث	**tha**	th	✓	✓	✓
ج	**gim**	j	✓	✓	✓
ح	**ha**	H	X	✓	✓
خ	**kha**	kh	✓	✓	✓
د	**da**	d	✓	✓	✓
ذ	**dal**	dh	✓	✓	✓
ر	**ra**	r	✓	✓	✓
ز	**zay**	z	✓	✓	✓
س	**sin**	s	✓	✓	✓
ش	**shin**	sh	✓	✓	✓
ص	**saad**	S	X	✓	✓
ض	**daad**	T	X	✓	✓
ط	**Ta**	D	X	✓	✓
ظ	**za**	Z	X	✓	✓
ع	**ayn**	`	✓	✓	✓
غ	**gayn**	gh	✓	✓	✓
ف	**Fa**	f	✓	✓	✓
ق	**qa**	q	✓	✓	✓

Qalam

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
ك	**ka**	k	✓	✓	✓
ل	**La**	l	✓	✓	✓
م	**mim**	m	✓	✓	✓
ن	**nun**	n	✓	✓	✓
ه	**ha**	h	✓	✓	✓
و	**wa**	w	✓	✓	✓
ى	**ya**	y	✓	✓	✓
ة	**ta marbuta**	h,t	✓	✓	✓
ء	**hamza**	'	✓	✓	✓

The one-to-one matches used by ISO 233-2 are present, as are a few two-English-letters to one-Arabic-character equivalents. And, rather than using non-native diacritics that an OPAC may not display, that ASCII will not recognize, and that few but linguistic professionals will understand, Qalam has put capital letters in their place. For example, ط is no longer ṭ, but instead T. As mentioned earlier, patrons and catalogers will almost certainly ignore diacritics they cannot read. A capitalization will, at the very least, alert the reader to the fact that that though a 'T' sound is involved here, it is not a direct match. Admittedly, on that front the benefit may be very slim, but at the very least the difference is expressed without compromising the OPAC or ASCII compatibility. However, on a negative note, the capitalization is not as phonetically accurate as a diacritic, costing Qalam five letters in the accuracy portion.

Qalam

Evaluation of:	Criteria	Possible points	Points scored
Accuracy	Is this transliterated character an accurate representation of the pronunciation of the vernacular?	60	50
Usability	Is this character ASCII compatible?	30	30
	Does this character use non-native diacritics?	30	30
Total points		120	110

SATTS

SATTS, the Standard Arabic Transliteration System, is a Latin Morse method most often employed by the military and communications companies (Standard, 2002). It is completely ASCII compatible. Attempts to date the creation and to identify the source of its origination remain unclear.

SATTS

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII	Non-use of Diacritics
ا	**alif**	A	✓	✓	✓
ب	**ba**	B	✓	✓	✓
ت	**Ta**	T	✓	✓	✓
ث	**tha**	C	X	✓	✓
ج	**gim**	J	✓	✓	✓
ح	**ha**	H	✓	✓	✓
خ	**kha**	O	X	✓	✓
د	**da**	D	✓	✓	✓
ذ	**dal**	Z	X	✓	✓
ر	**ra**	R	✓	✓	✓
ز	**zay**	;	X	✓	✓
س	**sin**	Z	X	✓	✓
ش	**shin**	:	X	✓	✓
ص	**saad**	X	X	✓	✓
ض	**daad**	V	X	✓	✓
ط	**Ta**	U	X	✓	✓
ظ	**za**	Y	X	✓	✓
ع	**ayn**	`	✓	✓	✓
غ	**gayn**	G	✓	✓	✓
ف	**Fa**	F	✓	✓	✓
ق	**qa**	Q	✓	✓	✓
ك	**ka**	K	✓	✓	✓
ل	**La**	L	✓	✓	✓
م	**mim**	M	✓	✓	✓

SATTS

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
ن	**nun**	N	✓	✓	✓
ه	**ha**	~	X	✓	✓
و	**wa**	W	✓	✓	✓
ى	**ya**	Y	✓	✓	✓
ة	**ta marbuta**	@	X	✓	✓
ء	**hamza**	E	X	✓	✓

Like Qalam, SATTS completely eliminated the diacritics that plague ISO 233-2 and ALA-LC. However, unlike Qalam, every single Arabic character is matched one-to-one with a key on a standard English keyboard. Thus, there is zero confusion as to which Arabic character writers intend to represent when they input text.

However, there are of course multiple phonetic problems in this situation. Though the characters used are completely compatible, a U does not represent the "*ta*" sound of ط and a colon does not represent the "*shin*" sound of a ش.

SATTS

Evaluation of:	Criteria	Possible points	Points scored
Accuracy	Is this transliterated character an accurate representation of the pronunciation of the vernacular?	60	34
Usability	Is this character ASCII compatible?	30	30
	Does this character use non-native diacritics?	30	30
Total points		120	94

Arabic Chat Alphabet (Arabesh)

The Arabic chat alphabet, which is known as "Arabesh" in some circles and as "Arabizi" in others, is the natural offshoot of native-Arab speakers using technological interfaces that do not, or at least once did not, support the Arabic vernacular (Arabic chat alphabet, 2005). Before the Mid-East adapted cell phone technology to be Arabic-enabled, much of the region was dependent upon devices that only supported English text. In addition, not all computer operating systems, especially a decade or so ago, supported Arabic vernacular, which inevitably led to a user-created transliteration system similar to what took place in China and Japan.

Arabic Chat Alphabet (Arabesh)

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII	Non-use of Diacritics
ا	**alif**	a	✓	✓	✓
ب	**ba**	b	✓	✓	✓
ت	**Ta**	t	✓	✓	✓
ث	**tha**	s/th	✓	✓	✓
ج	**gim**	g/j	✓	✓	✓
ح	**ha**	7	X	✓	✓
خ	**kha**	5/kh	✓	✓	✓
د	**da**	d	✓	✓	✓
ذ	**dal**	z	X	✓	✓
ر	**ra**	r	✓	✓	✓
ز	**zay**	z	✓	✓	✓
س	**sin**	s	✓	✓	✓
ش	**shin**	sh	✓	✓	✓
ص	**saad**	S/9	X	✓	✓
ض	**daad**	D/9'	X	✓	✓
ط	**Ta**	TH/T/6	X	✓	✓
ظ	**za**	Z/TH/6'	✓	✓	✓
ع	**ayn**	3	X	✓	✓

Arabic Chat Alphabet (Arabesh)

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
غ	**gayn**	gh/3'	✓	✓	✓
ف	**Fa**	f/ph	✓	✓	✓
ق	**qa**	q/8/9	✓	✓	✓
ك	**ka**	k	✓	✓	✓
ل	**La**	l	✓	✓	✓
م	**mim**	m	✓	✓	✓
ن	**nun**	n	✓	✓	✓
ه	**ha**	h	✓	✓	✓
و	**wa**	w	✓	✓	✓
ى	**ya**	i/y	X	✓	✓
ة	**ta marbuta**	h,t	✓	✓	✓
ء	**hamza**	2	X	✓	✓

As this author can attest, texting was alive and well in Dubai in 2003, and more often than not the messages high school and college age students were exchanging were in either English or Arabesh. Such a phenomenon is not without detractors, or at least investigators, as a documentary, Arabizi, produced by Dalia al-Kury, looked at the phenomenon in Amman, Jordan in 2005 (Ejeilat, 2005). Thus, no matter what score the system receives here, a good case could be made for a system that is now the favorite of Arab youth.

Obviously, to utilize the tools above, Arabesh uses no diacritics or any non-ASCII compatible letters. However, it is prone to phonetic inaccuracies, as users sometimes substitute numbers in the place of letters that have no direct English equivalent.

A problem in the evaluation of this aspect is that the language is not set. It varies from user to user and some letters can have as many as four different characters

representing it. Thus, the author concluded that using what is most likely to be the most common form of the language (ascertained through personal experience in the Mid-East) to score its accuracy component would be the most prudent course of action. However, since there are no statistics on such matters, this portion of the Arabic chat alphabet system's score is at least somewhat subjective.

Arabic Chat Alphabet (Arabesh)

Evaluation of:	Criteria	Possible points	Points scored
Accuracy	Is this transliterated character an accurate representation of the pronunciation of the vernacular?	60	44
Usability	Is this character ASCII compatible?	30	30
	Does this character use non-native diacritics?	30	30
Total points		120	104

Buckwalter

The Buckwalter transliteration method, developed at Xerox in 1990, "is used for representing exact orthographical strings of Arabic in email and other environments where the display of real Arabic script is impractical or impossible" (Buckwalter, 2001; Arabic, 2008). This system is ASCII compatible and, like Qalam and Arabesh, does not use diacritics.

Buckwalter

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII	Non-use of Diacritics
ا	**alif**	A	✓	✓	✓
ب	**ba**	B	✓	✓	✓
ت	**Ta**	T	✓	✓	✓
ث	**tha**	V	X	✓	✓
ج	**gim**	J	✓	✓	✓
ح	**ha**	H	✓	✓	✓
خ	**kha**	X	X	✓	✓
د	**da**	D	✓	✓	✓
ذ	**dal**	*	X	✓	✓
ر	**ra**	R	✓	✓	✓
ز	**zay**	Z	✓	✓	✓
س	**sin**	S	✓	✓	✓
ش	**shin**	$	X	✓	✓
ص	**saad**	S	X	✓	✓
ض	**daad**	D	X	✓	✓
ط	**Ta**	T	X	✓	✓
ظ	**za**	Z	✓	✓	✓
ع	**ayn**	E	X	✓	✓
غ	**gayn**	G	✓	✓	✓
ف	**Fa**	F	✓	✓	✓
ق	**qa**	Q	✓	✓	✓
ك	**ka**	K	✓	✓	✓
ل	**La**	L	✓	✓	✓

Buckwalter

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
م	**mim**	M	✓	✓	✓
ن	**nun**	N	✓	✓	✓
ه	**ha**	H	✓	✓	✓
و	**wa**	W	✓	✓	✓
ى	**ya**	Y	✓	✓	✓
ة	**ta marbuta**	P	X	✓	✓
ء	**hamza**	‘	✓	✓	✓

The drawback to the Buckwalter method is that it incorporates several instances where the character used to match an Arabic letter is nothing close to the original pronunciation. Buckwalter does not do this as often as SATTS, but it is a negative nonetheless. In addition, much like Qalam, Buckwalter uses capitalization to emphasize the presence of a different pronunciation from another Arabic character that uses the same Roman letter, only in lower-case form. While this helps a user differentiate one from another, complete phonetic accuracy requires a diacritic.

Buckwalter

Evaluation of:	Criteria	Possible points	Points scored
Accuracy	Is this transliterated character an accurate representation of the pronunciation of the vernacular?	60	42
Usability	Is this character ASCII compatible?	30	30
	Does this character use non-native diacritics?	30	30
Total points		120	102

ALA-LC/UNGEGN

The transliteration system for the American Library Association and the Library of Congress, which is also used by the Online Computer Library Center, differs from the system created by the United Nations Experts on Groups of Geographical Names only in terms of the shape of the diacritical mark placed beneath five letters. For the ض ص ح ط and ظ there is a dot underneath in ALA-LC and a curved line, used in regular Turkish, for UNEGN. Because there is only this superficial distance, at least for basic character transliteration, the systems have been combined.

ALA-LC/UNGEGN

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII	Non-use of Diacritics
ا	**alif**	(omit)	X	✓	✓
ب	**ba**	b	✓	✓	✓
ت	**ta**	t	✓	✓	✓
ث	**tha**	th	✓	✓	✓
ج	**gim**	j	✓	✓	✓
ح	**ha**	ḥ	✓	X	X
خ	**kha**	kh	✓	✓	✓
د	**da**	d	✓	✓	✓
ذ	**dal**	dh	✓	✓	✓
ر	**ra**	r	✓	✓	✓
ز	**zay**	z	✓	✓	✓
س	**sin**	s	✓	✓	✓
ش	**shin**	sh	✓	✓	✓
ص	**saad**	ṣ	✓	X	X
ض	**daad**	ḍ	✓	X	X
ط	**ta**	ṭ	✓	X	X
ظ	**za**	ẓ	✓	X	X
ع	**ayn**	`	✓	✓	✓
غ	**gayn**	gh	✓	✓	✓

ALA-LC/UNGEGN

Arabic Letter	Name of Letter	Transliteration	Phonetic Accuracy	ASCII Compatibility	Non-use of Diacritics
ف	**fa**	f	✓	✓	✓
ق	**qa**	q	✓	✓	✓
ك	**ka**	k	✓	✓	✓
ل	**la**	l	✓	✓	✓
م	**mim**	m	✓	✓	✓
ن	**nun**	n	✓	✓	✓
ه	**ha**	h	✓	✓	✓
و	**wa**	w	✓	✓	✓
ى	**ya**	y	✓	✓	✓
ة	**ta marbuta**	h,t	✓	✓	✓
ء	**hamza**	□ [1]	✓	X	✓

As stated from the beginning of this paper, ALA-LC/UNGEGN uses non-native diacritics and is not ASCII compatible. Though this system does both far less than ISO 233-2, these marks against it, six non-native, and five not compatible, harm what would otherwise be a useful system. This method does not shy away from using two letters to convey an Arabic sound when needed, and it does a solid job of making sure no two characters could be mistaken for one another after transliteration.

There is one odd anomaly, however, as the *alif*, (ا) the most common letter in Arabic, is omitted. It is simply not printed if it stands alone or modifies a consonant. The *alif* is only printed when it is being modified by another long vowel or when it is supporting a *hamza*, **ء**,(Barry, 1997).

[1] The ALA-LC/UNGEGN character for the hamza, which resembles an apostrophe, will not display in MSWord

That said, the rest of the system is very accurate, phonetically, but its insistence on diacritics native to neither language and that cannot be found anywhere within ASCII's 255-character set, means that it simply cannot be effectively utilized in computer code or by many universities' OPACs, catalogers, and patrons.

ALA-LC/UNGEGN

Evaluation of:	Criteria	Possible points	Points scored
Accuracy	Is this transliterated character an accurate representation of the pronunciation of the vernacular?	60	58
Usability	Is this character ASCII compatible?	30	24
	Does this character use non-native diacritics?	30	25
Total points		120	107

Korean Scoring

And, to show how the Korean systems would have been scored:

McCune-Reischaeur

Evaluation of:	Criteria	Possible	Points scored
Accuracy	Is this transliterated character an accurate representation of the pronunciation of the vernacular?	66	56
Usability	Is this character ASCII compatible?	33	29
	Does this character use non-native diacritics?	33	29
Total points		132	114

Revised Romanization

Evaluation of:	Criteria	Possible points	Points scored
Accuracy	Is this transliterated character an accurate representation of the pronunciation of the vernacular?	66	66
Usability	Is this character ASCII compatible?	33	33
	Does this character use non-native diacritics?	33	33
Total points		132	132

CONCLUSION

Accuracy/Usability Scores

	ISO 233-2	Qalam	SATTS	Arabesh	Buckwalter	ALA/LC-UNGEGN
Accuracy	58	50	34	44	42	58
Usability	36	60	60	60	60	49
Total	94	110	94	104	102	107

System Ranking Scores

Rank	System	Score
1	Qalam	110
2	ALA/LC-UNGEGN	107
3	Arabesh	104
4	Buckwalter	102
5	SATTS	94
5	ISO 233-2	94

As the numbers above show, Qalam emerges as the winner, though ALA-LC claims a very respectable second. Nevertheless, the recommendation made here is that the American Library Association, the Library of Congress, the United Nations, and the Online Computer Library Center should abandon their own transliteration method, one that uses non-ASCII characters native to neither Arabic nor English, in favor of Qalam, which avoids all diacritics while maintaining complete ASCII-compatibility and almost the same level of phonetic accuracy.

By making such a move, ALA-LC, OCLC, and the UN will be giving non-linguists the best access to information written in, or accessed by, transliterated Arabic.

The characters are familiar to all English speakers and the letters and symbols can be displayed on any OPAC. In order to make access to Arabic materials easier for people who may not be experts in the language, this is a positive step.

Also noteworthy, as the Accuracy/Usability Scores show, achieving high marks in both usability and accuracy is a near impossibility. The differences between English and Arabic are simply too great to make any one transliteration system a seamless transition from one to the other. Perhaps for linguists, ISO 233-2 or ALA-LC may be the most useful, but for American libraries, catalogers, and patrons, neither fits as well as an ASCII compatible system with no diacritics and a very, though not perfectly, accurate phonetic transliteration.

Finally, seeing that Arabesh, created more or less by Arab teenagers texting and IM'ing one another, beats out three scientifically engineered systems lends credence to the statement by T.E. Lawrence (Lawrence of Arabia),

> Arabic names won't go into English exactly, for their consonants are not the same as ours, and their vowels, like ours, vary from district to district. There are some 'scientific systems' of transliteration, helpful to people who know enough Arabic not to need helping, but a washout for the world. I spell my names anyhow, to show what rot the systems are. (Whitaker, 2002)

AREAS FOR FURTHER STUDY

A survey of Arabic catalogers in both Anglophone and Arab countries could be conducted. The catalogers could evaluate the transliteration systems according to the same parameters used in this investigation. The study should be controlled for levels of Arabic ability in order to gauge whether or not there is a correlation between fluency and transliteration preference.

The same study could be conducted with library patrons, computational linguists, military translators, Arabic studies professors, and other professionals with a vested interest in this field.

OPACs of every major operating system, including more than just the most recent version, could be surveyed to see which systems support which diacritics.

APPENDICES

Appendix A: Arabic Transliteration Methods

Arabic Letter	Name of Letter	ISO 233-2	Qalam	SATTS	Arabesh	Buckwalter	ALA/LC-UNGEGN
ب	ba	b	B	B	b	b	b
ت	ta	t	T	T	t	t	t
ث	tha	t̲	Th	C	s/th	v	th
ج	gim	ǧ	J	J	g/j	j	j
ح	ha	ḥ	H	H	7	H	ḥ
خ	kha	ḫ	Kh	O	5/kh	x	kh
د	da	d	D	D	d	d	d
ذ	dal	d̲	Dh	Z	z	*	dh
ر	ra	r	R	R	r	r	r
ز	zay	z	Z	;	z	z	z
س	sin	s	S	Z	s	s	s
ش	shin	š	Sh	:	sh	$	sh
ص	saad	ṣ	S	X	S/9	S	ṣ
ض	daad	ḍ	T	V	D/9'	D	ḍ
ط	ta	ṭ	D	U	TH/T/6	T	ṭ
ظ	za	z̲	Z	Y	Z/TH/6'	Z	z̲
ع	ayn	ʿ	`	`	3	E	ʿ
غ	gayn	ḡ	Gh	G	gh/3'	g	gh
ف	fa	f	F	F	f/ph	f	f
ق	qa	q	Q	Q	q/8/9	q	q
ك	ka	k	K	K	k	k	k
ل	la	l	L	L	l	l	l
م	mim	m	M	M	m	m	m
ن	nun	n	N	N	n	n	n
ه	ha	h	H	~	h	h	h
و	wa	w	W	W	w	w	w
ى	ya	y	Y	Y	i/y	y	y
ة	ta marbuta	h,t	h,t	@	h,t	p	h,t
ء	hamza	ʻ	ʻ	E	2	ʻ	ʻ

Appendix B: ASCII Charts

These are the ASCII codes as used in Microsoft Excel.

ASCII non-printing control characters 0-31	
Decimal	**Character**
0	null
1	start of heading
2	start of text
3	end of text
4	end of transmission
5	inquiry
6	acknowledge
7	bell
8	backspace
9	horizontal tab
10	line feed/new line
11	vertical tab
12	form feed/new page
13	carriage return
14	shift out
15	shift in
16	data link escape
17	device control 1
18	device control 2
19	device control 3
20	device control 4
21	negative acknowledge
22	synchronous idle
23	end of transmission block
24	cancel
25	end of medium
26	substitute
27	escape
28	file separator
29	group separator
30	record separator
31	unit separator

ASCII printing characters 32-127	
Decimal	**Character**
32	space
33	!
34	"
35	#
36	$
37	%
38	&
39	'
40	(
41	)
42	*
43	+
44	,
45	-
46	.
47	/
48	0
49	1
50	2
51	3
52	4
53	5
54	6
55	7
56	8
57	9
58	:
59	;
60	<
61	=
62	>
63	?
64	@
65	A
66	B
67	C
68	D
69	E
70	F
71	G
72	H

ASCII printing characters 32-127	
Decimal	**Character**
73	I
74	J
75	K
76	L
77	M
78	N
79	O
80	P
81	Q
82	R
83	S
84	T
85	U
86	V
87	w
88	X
89	Y
90	Z
91	[
92	\
93	]
94	^
95	_
96	`
97	a
98	b
99	c
100	d
101	e
102	f
103	g
104	h
105	i
106	j
107	k
108	l
109	m
110	n
111	o
112	p
113	q
114	r

ASCII printing characters 32-127	
Decimal	**Character**
115	s
116	t
117	u
118	v
119	w
120	x
121	y
122	z
123	{
124	|
125	}
126	~
127	DEL

Extended ASCII printing characters 128-223	
Decimal	**Character**
128	Ç
129	ü
130	é
131	â
132	ä
133	à
134	å
135	ç
136	ê
137	ë
138	è
139	ï
140	î
141	ì
142	Ä
143	Å
144	É
145	æ
146	Æ
147	ô
148	ö
149	ò
150	û
151	ù
152	ÿ
153	Ö
154	Ü
155	¢
156	£
157	¥
158	₧
159	ƒ
160	á
161	í
162	ó
163	ú
164	ñ
165	Ñ
166	ª
167	º
168	¿

Extended ASCII printing characters 128-223	
Decimal	**Character**
169	⌐
170	¬
171	½
172	¼
173	¡
174	«
175	»
176	▒
177	▒
178	▓
179	│
180	┤
181	╡
182	╢
183	╖
184	╕
185	╣
186	║
187	╗
188	╝
189	╜
190	╛
191	┐
192	└
193	┴
194	┬
195	├
196	─
197	┼
198	╞
199	╟
200	╚
201	╔
202	╩
203	╦
204	╠
205	═
206	╬
207	╧
208	╨
209	╤
210	╥

Extended ASCII printing characters 128-223	
Decimal	**Character**
211	╙
212	Ô
213	╒
214	╓
215	╫
216	╪
217	┘
218	┌
219	█
220	▄
221	▌
222	▐
223	▀
224	α
225	ß
226	Γ
227	π
228	Σ
229	σ
230	µ
231	τ
232	Φ
233	Θ
234	Ω
235	δ
236	∞
237	φ
238	ε
239	∩
240	≡
241	±
242	≥
243	≤
244	⌠
245	⌡
246	÷
247	≈
248	≈
249	·
250	·
251	√
252	ⁿ

Extended ASCII printing characters 128-223	
Decimal	**Character**
253	²
254	■
255	

Appendix C: The End of an Arabic Text

This Arabic text, in full diacritics, is often found on the last page of an Arabic book. The words mean "Completed by the grace of God." (Wilson, 2004)

BIBLIOGRAPHY

Arabic Chat Alphabet. (2005). *Wikipedia, the free encyclopedia.* Retrieved March 7, 2008 from http://en.wikipedia.org/wiki/Arabic_chat_alphabet

Arabic Transliteration/Encoding Chart. (2008). Retrieved from Xerox website March 26, 2008, from http://www.xrce.xerox.com/competencies/content-analysis/arabic/info/translit-chart.html

Barry, R.K. (1997). *ALA-LC Romanization tables: Transliteration schemes for non-Roman scripts*. Washington: Cataloging Distribution Service, Library of Congress. Retrieved March 7, 2008, from http://www.loc.gov/catdir/cpso/romanization/arabic.pdf

Beesley, K. (1998). *Romanization, transcription, and transliteration*. Retrieved from Xerox website March 26, 2008, from http://www.xrce.xerox.com/competencies/content-analysis/arabic/info/romanization.html

Buckwalter, T. (2001). *Buckwalter Arabic transliteration.* Retrieved March 7, 2008 from, http://www.qamus.org/transliteration.htm

Eilts, J. (1995). Non-Roman script materials in North American libraries: automation and international exchange. *International Federation of Library Associations and Institutions,* 61st general conference. Retrieved March 19, 2008, from http://www.ifla.org/IV/ifla61/61-eilj.htm

Ejeilat, L. (2005). 'Arabizi'… the phenomenon in Amman??? *Lina's Turmoil Blog.* Retrieved March 26, 2008, 2008, from http://linasturmoil.blogspot.com/2005_07_01_archive.html

Heddaya, A. (1985). *Qalam: A convention for morphological Arabic-Latin-Arabic transliteration.* Retrieved March 7, 2008, from http://langs.eserver.org/qalam.txt

International Standards Organization. (2008). *ISO – Standards development processes*. Retrieved March 26, 2008, from http://www.iso.org/iso/standards_development/processes_and_procedures/how_are_standards_developed.htm

ISO 233. (2005). *Wikipedia, the free encyclopedia.* Retrieved March 7, 2008, from http://en.wikipedia.org/wiki/ISO_233

Ministry of Culture & Tourism. (2000). *Romanization of Korean*. Retrieved March 26, 2008, from http://www.korea.net/korea/kor_loca.asp?code=A020303

Standard Arabic Technical Transliteration System. (2002). *Wikipedia, the free encyclopedia.* Retrieved March 7, 2008, from http://en.wikipedia.org/wiki/Standard_Arabic_Technical_Transliteration_System

UNGEGN Working Group on Romanization Systems. (2003). Arabic, *Report on the current status of United Nations Romanization systems for geographical names, Version 2.2.* Retrieved March 17, 2008, from http://www.eki.ee/wgrs/rom1_ar.htm

UNGEGN Working Group on Romanization Systems. (2003). Korean, *Report on the current status of United Nations Romanization systems for geographical names, Version 2.2.* Retrieved March 26, 2008, from http://www.eki.ee/wgrs/rom2_ko.htm

Variations in the spelling of Qadhafi. (2004). *Find out more about Libya.* Retrieved March 16, 2008, from http://www.geocities.com/Athens/8744/spelling.htm

Whitaker, B. (2002). Lost in translation. *The Guardian.* Manchester: Guardian News and Media Limited. Retrieved March 16, 2008, from http://www.guardian.co.uk/world/2002/jun/10/israel1

Wilson, K. (2004). *A guide to copy cataloging Arabic materials.* MS thesis. University of North Carolina. Retrieved February 19, 2008 from http://ils.unc.edu/MSpapers/3092.pdf