

コンピュータの『国際化』に関する一考察

Beshr B. Bakhit, 遠山 元道, 浦 昭二

慶応義塾大学、管理工学研究科

『空腹の男に魚をやれば、彼はその日は腹を満たすことができる。
しかし、魚のつり方を教えてやれば、一生はらを満たします。』

(中国のことわざ)

概要 :

今日、社会が高度な工業化を達成するためには、何らかの形でのコンピュータを含む近代的なテクノロジーの導入が必要である。高度な工業化を達成するためには、主に2つのものが必要であると思われる。それは、技術的ノウハウとそれを実現するための経済的な資源である。ここにおいて、多くの発展途上国は重大な問題に直面していると言える。これらの国々は、一方では産油諸国のように経済的な資源を持ちながらも、ノウハウに欠き、他方では東南アジア諸国のように経済的な資源とノウハウの両方を欠くという困難な問題を抱えている。結果として、先進工業国と発展途上国の間の格差は毎年広がるばかりである。コンピュータ専門家の見地より、本論文ではサウジアラビアにおける『コンピュータのアラビア化』をケース・スタディとして取り上げ、この問題の技術的側面に考察を加えることにする。

序論 :

過去20年あまり、コンピュータは多くの分野（産業、金融、...）においてその有効性を発揮し、よって今日、高度な工業化を目指す国にとって、それは必要不可欠なものとなった。例えば、業務の取り扱いをオンラインではなくマニュアルで行っている銀行があるでしょう。顧客数があるレベルを越えると、業務の取り扱いをマニュアルで行うことは極めて困難になり、オートメーション化が必要となる。さもないければこの銀行は他の『ハイテク』銀行に顧客を奪われてしまうであろう。

さらに強調しておかなければいけないのは、これらのコンピュータが主に欧米諸国で開発されたということである。その為、英語用に開発されたコンピュータがドイツ語やフランス語を処理するには、若干の修正が必要となる。日本は、これらのコンピュータが日本語に適用できるよう独自のコンピュータ・システムを再開発した代表的な国であると言える。したがって、母国語を処理することが可能なコンピュータの開発を望む多くの国々は、日本を手本としているのである。なぜなら、従来のシステムを用いても、それが利用者の母国語を処理することができないのであれば、それは役に立たないからである。例えば、漢字または仮名を処理することができない DBMS は日本人顧客データベースの構築に、さほど役に立たないであろう。

コンピュータのアラビア化について :

北部アフリカおよび中東諸国において使用されている言語は、アラビア語である。さらに、ペルシャ語（イラン）、ウルドゥー語（パキスタン）、マレー語（マレーシア）およびウイグル語（中国の新疆部）では、それぞれの言語を書くためにアラビア文字を用いている。これは、これらの国々がイスラム教国であるからである。したがって、これらの国の国民が一度も学んだことのない言語を

『読む』ことは可能であるが、その意味を理解することはできないのである。中国語の文章の漢字の意味がわかって、それを読むことができない日本人に対比して考えることが可能であろう。

では、ここでコンピュータ上の問題について検討しよう。

まず、アラビア文字について検討する。アラビア文字は、日本語の仮名がローマ字で表わせる(図1)のと同じように、英字で表わせないものだろうか。もしこれが可能ならば、最大の難問であるデータ入力为解决し、残る問題はアラビア文字の活字(Print Font)を開発することだけである。ここで、アラビア語の二つの単語をとりあげてみよう。『شركة』(仕事)と『شركة』(希望)というこのふたつの単語は、両方とも英語では(A M A L)と書く。つまり、すべてのアラビア文字、そしてそれらの音が必ずしも英字で正確に表わせるとは限らないのである。上記の例の場合、『ر』と『ا』は両方とも(A)で表わされる。アラビア字と英字は、Many-To-Many Mappingを構成しているのである(図2)。よって、我々の最初の問いに対する答えは、『否』である。アラビア文字独自の文字セットを開発しなければいけないのである。

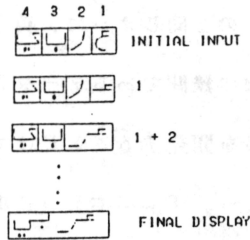
ASMO-449 Code Set(図3)は、正式に採用されているアラビア文字コードである。これは、すべてのアラビア文字と母音添加記号、およびCRやHTなどのコントロール文字を含むものである。したがって、このコードを用いることによって、すべてのアラビア語データを十分かつ正確に表わすことが可能となる。

ところで、アラビア語は以下の特徴を有し、コンピュータ専門家の注意を必要とする。

- 1- アラビア文字は、36の文字によって表わされる。
- 2- 8つ母音記号(添加記号)がある。
- 3- アラビア数字(注1)は左から右へ書かれ、読み方も同様である。

注1: 実際、アラブ諸国で使われている数字は、いわゆる『アラビア数字』とは異なる(図3を参照せよ)。

- 4- アラビア文字は、右から左へ書かれ、読み方も同様である。
- 5- 単語を構成する文字は、その単語における位置によって互いに結合する。これは、レキシカル分析ルーティーン (Lexical Analysis Routine) を用いて、プリント/ディスプレイ・バッファーにおいて行われる (例1)。



例1 アラビア文字の結合

サウジアラビアにおけるアラビア化の努力:

コンピュータのアラビア化の分野は、1970年代の中頃まで注目を浴びることがなかった。教育や医療といったもっと国益に関わる重要な分野に、資源と努力が向けられていたからである。オイル・マネーによる収入がコンピュータのアラビア化の道を開いた。しかし、残念なことに、これらの努力は時期的に遅かったため、政府、民間、および教育のセクターにおいて、アラビア化を実現するためには、それだけ多くの努力と資源を必要とすることになったのである。

コンピュータのアラビア化に向けての第一歩は、標準文字コード・セットを決めることであった。そのために、1980年6月には、リヤドにおいて『コンピュータにおけるアラビア語のための文字セットおよびキーボードの標準化に関する国際シンポジウム』が開催され、その結果、ASMO-449セットが正式に採用されることになった。(キーボードに関しては、何も採用されなかった)。

有効なキーボード・レイアウトを作成する上で、1文字および2文字の度数の分析が必要である。多くの研究者が、最適キーボード作成のための綿密な研究を行い、これらの研究結果からは異なるレイアウト[1]、[2]、[3]が提案・作成されている。

アラビア語における文字の出現頻度は、Text-Dependable である。例えば、アラビア名、アラビア語辞書またはアラビア語テキストは、異なる度数を示す。アラビア語圏でも、国によっては度数が異なる。

結論から言えば、現在、多くの『最適』と呼ばれるキーボードが市場に出回っているが、標準キーボードが利用者のために開発されなければいけない。

標準キーボードが開発されればよいのであって、どのレイアウトが正式に採用されるかは、さほど問題ではない。なぜなら、QWERTY キーボードはかなり不適当なキーボードであるにもかかわらず、誰も苦情を言っていないからである。では、ここでその応用面についてみてみよう。

サウジアラビアの大学における学問上の研究のほとんどが、『SAUDIA』や『KOWARAZMI』などを始めとするアラビア語のプログラミング言語を開発することに専念してきたということは注目に値する。これらのプログラミング言語は、商業用ではなく、英語を用いないプログラミングを高校生および大学生に教えるために、主に開発された。これが開発されたもう1つの目的というのは、英語で考えて、アラビア語でプログラミングする際に生じる混乱を避けるためである。(既に説明した英語とアラビア語の文構造の違いを参照されたい)。

コンピュータのアラビア化に伴う技術的問題：

先にも触れたように、すべてのアラビア文字を7ビットで表わすASMO-449が標準アラビア文字コード・セットとして採用された。しかし、現在使われているソ

ソフトウェアは (EBCDIC 等) 用に作成されたものがほとんどで、アラビア語データでこれらのソフトウェアを用いる際、多くの問題が生じた。さらに、ASMO-449にはで英文字が含まれていないので、英語のデータを処理するには、他の文字コードセット (ASCII, EBCDIC等) への切替えを指示する制御シフト・アウト/シフト・イン (SO/SI)の動作を行なう必要があり、これは極めて不便であることが判明した。實際上、1つのデータセット中がこの切替えを行なうことはきわめて困難なので、例えばアラビアの銀行では、英字名の顧客とアラビア字名の顧客は別のデータベースに登録される。最近では、8ビット・コード (ASMO DS-708 図4) を採用する傾向にあるが、このコードが抱える問題は、文字が英語であるかアラビア語であるか識別し、それに沿ってディスプレイ/プリントするソフトウェアがもう1つ必要であるということである。

ソフトウェアについて言えば、アラビア語の本質が反映されているソフトウェアというのは大変限られており、市場に出回っているほとんどのソフトウェアは、英語用の『アラビア版』である。この分野での、努力がさらに必要であると思われる。今日に至っても大型コンピュータにアラビア語 OS が実施された例は全くない。(それに似たものは既にあるが...)

コンピュータのアラビア化を拒んでいるもう1つの問題は、著作権の問題である。ソフトウェアの違法コピーは多くの製造者にとって深刻な問題であり、これに対処するために罰金の重い著作権法などの制定が検討されている。なお、サウジアラビアにおいては、違法コピーそれ自体よりも著作権法の意識が希薄であるということが、ソフトウェア製造者の心配事となっている。したがって、多くの会社はソフトウェアに対する著作権法が制定されない限りは、コンピュータのアラビア化を行なわないという立場をとるに至っている。現在、政府関連機関がそのような著作権法の制定に向けて動き出しており、その実施は時間の問題である。

これからの課題：

近年、『国際人』ということばがしばしば聞かれる。コンピュータの専門家は、いかにして自分の分野で『国際人』になれるであろうか。この問いに対する一つの答えは、まず、コンピュータ産業に関しては、その産業に携わる1人1人の人間がコンピュータの分野における発展途上国と日本の間には存在する格差を縮めるために、自分の会社はどのような努力を行なっているか、また、英語ないしは日本語以外の言語で自分のところのシステムを使用することは可能か、と考えることだと思う。どの国も、その母国語によって機能が可能なコンピュータ・システムを持つ権利を有しており、利用に際して英語又は母国語以外のことばが必要であるというのは問題がある。

次に教育の分野においては、発展途上国からの学生にどれだけ教育の機会を提供し、彼らが自国の工業化に貢献することができるような手をさしのべたか、という問いを行なうことが重要であると思われる。さらに、研究者の立場から、発展途上国が直面するコンピュータの問題についてどれだけ研究し、その解決法を見出そうとしたか、問うことが大切である。

これらの問題について考え検討することが『コンピュータ国際人』への道である。

				0	1	2	3	4	5	6	7	
0	0	0	0	0	NUL	TC	SP	•	ا	ذ	-	...
0	0	0	1	1	TC	DC	!	۱	۰	ر	ف	...
0	0	1	0	2	TC	DC	"	۲	۳	ز	...	
0	0	1	1	3	TC	DC	#	۳	۱	س	ع	
0	1	0	0	4	TC	DC	¤	۴	ف	ا	ع	
0	1	0	1	5	TC	TC	%	۵	!	ص	ع	
0	1	1	0	6	TC	TC	&	۶	ذ	ذ	ع	
0	1	1	1	7	DEL	TC	'	۷	ا	ط	ع	
1	0	0	0	8	FE	CAN)	۸	ب	ظ	ع	
1	0	0	1	9	FE	EM	(۹	ة	ع	ع	
1	0	1	0	10	FE	SUB	*	:	ت	ع	ع	
1	0	1	1	11	FE	ESC	+	;	ع	ع	ع	
1	1	0	0	12	FE	IS	.	۱۲	۱	ع	ع	
1	1	0	1	13	FE	IS	-	=	>	ع	ع	
1	1	1	0	14	SO	IS	.	<	۸	ع	ع	
1	1	1	1	15	SI	IS	/	?	د	-	DEL	

図3 ASMO-449

				0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	00	sp	0	a	P	'	D	WSP								
0	0	0	1	01	!	1	A	Q	a	q									
0	0	1	0	02	"	2	B	R	b	r									
0	0	1	1	03	#	3	C	S	c	s									
0	1	0	0	04	\$	4	D	T	d	t									
0	1	0	1	05	%	5	E	U	e	u									
0	1	1	0	06	&	6	F	V	f	v									
0	1	1	1	07	'	7	G	W	g	w									
1	0	0	0	08	(8	H	X	h	x									
1	0	0	1	09)	9	I	Y	i	y									
1	0	1	0	10	*	:	J	Z	j	z									
1	0	1	1	11	+	;	K	C	k	c									
1	1	0	0	12	,	.	L	\	l	l									
1	1	0	1	13	-	=	M	J	m	j									
1	1	1	0	14	.	>	N	^	n	^									
1	1	1	1	15	/	?	0	_	o	_									

図4 ASMO DS-708

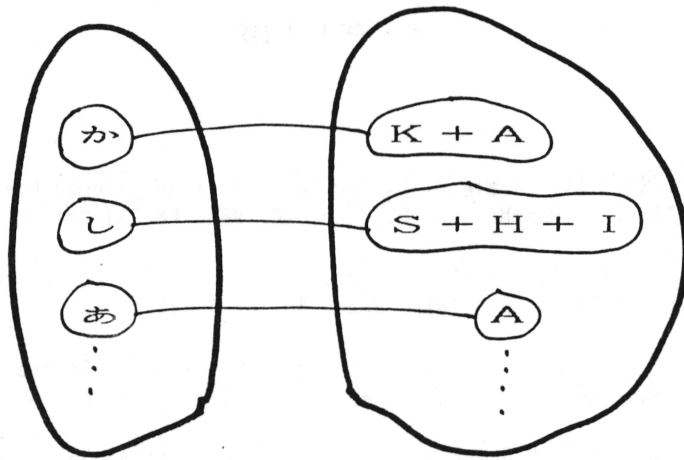


図 1 日本語を英語で表わせる

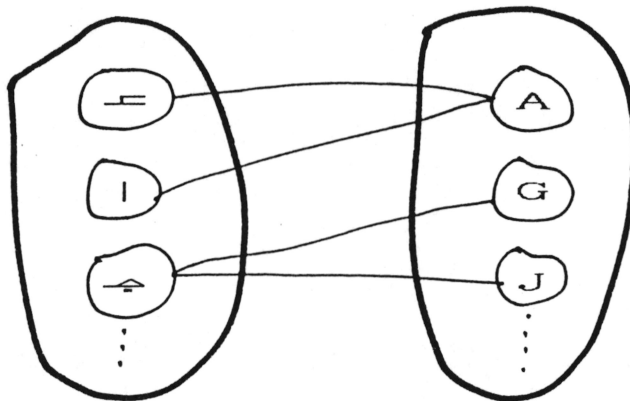


図 2 アラビア語を英語で表わせる

参考図書

- 1- A.J. AL-Khalili, "An algorithm For An Intelligent Arabic Computer Terminal," Int. J. Man-Machine Studies ,20,1984, PP.331-341.
- 2- G. Kubba, "The Impact Of Computers On Arabic Writing, Character Processing And Teaching," Info. Proc. 80, PP.36-40.
- 3- S.Hyder, K. Ibrahim, "Optimal Placement Of Isolated Characters On Arabic Keyboards," Proc. Int. Symp. Std., Riyadh, June 1980 PP.961-65.
- 4- Arab League : Arab Organization for Standardization and Metrology, " Arab Standard Spec. 449," ASMO 1982.
- 5- Saudi Arabian Standards Organization, "Data Processing 7-Bit Coded Arabic Character Set for Information Interchange," SASO 1983.

本 PDF ファイルは 1987 年発行の「第 28 回プログラミング・シンポジウム報告集」をスキャンし、項目ごとに整理して、情報処理学会電子図書館「情報学広場」に掲載するものです。

この出版物は情報処理学会への著作権譲渡がなされていませんが、情報処理学会公式 Web サイトに、下記「過去のプログラミング・シンポジウム報告集の利用許諾について」を掲載し、権利者の検索をおこないました。そのうえで同意をいただいたもの、お申し出のなかったものを掲載しています。

https://www.ipsj.or.jp/topics/Past_reports.html

過去のプログラミング・シンポジウム報告集の利用許諾について

情報処理学会発行の出版物著作権は平成 12 年から情報処理学会著作権規程に従い、学会に帰属することになっています。

プログラミング・シンポジウムの報告集は、情報処理学会と設立の事情が異なるため、この改訂がシンポジウム内部で徹底しておらず、情報処理学会の他の出版物が情報学広場 (=情報処理学会電子図書館) で公開されているにも拘らず、古い報告集には公開されていないものが少からずありました。

プログラミング・シンポジウムは昭和 59 年に情報処理学会の一部門になりましたが、それ以前の報告集も含め、この度学会の他の出版物と同様の扱いにしたいと考えます。過去のすべての報告集の論文について、著作権者（論文を執筆された故人の相続人）を探し出して利用許諾に関する同意を頂くことは困難ですので、一定期間の権利者搜索の努力をしたうえで、著作権者が見つからない場合も論文を情報学広場に掲載させていただきたいと思います。その後、著作権者が発見され、情報学広場への掲載の継続に同意が得られなかった場合には、当該論文については、掲載を停止致します。

この措置にご意見のある方は、プログラミング・シンポジウムの辻尚史運営委員長 (tsuji@math.s.chiba-u.ac.jp) までお申し出ください。

加えて、著作権者について情報をお持ちの方は事務局まで情報をお寄せくださいますようお願い申し上げます。

期間： 2020 年 12 月 18 日 ~ 2021 年 3 月 19 日

掲載日： 2020 年 12 月 18 日

プログラミング・シンポジウム委員会

情報処理学会著作権規程

<https://www.ipsj.or.jp/copyright/ronbun/copyright.html>