

LANGUAGE AND THE UNIFORMITY OF SPELLING*

ASMAH HAJI OMAR
University of Malaya
(Received September 14, 1971)

In a non-literate society the people involved in the use of language consist of two categories: the category of the speaker and that of the listener. When society is literate, two other categories are added to the ones already existing and these are the categories of the writer and the reader.

A literate society is a society which makes use of a writing system as a visual representation of its language code. A writing system can consist of a logogram, an ideogram, a syllabary, or an alphabet. A logogram is a system which uses separate symbols to represent individual words. The kind of writing we use for numerals is a logogram. The graphic symbol which is the Arabic numerical sign 3 stands for the word pronounced [θ ri:] (spelled *three*) in English, and /tiga/ (spelled *tiga*) in Malay. An ideogram is a system which uses a sign to express an "idea." Chinese writing system has often been cited as an instance of this system of writing. The third type of writing system, the syllabary, makes use of a separate graphic symbol to represent a syllable, while the fourth system, the alphabet system, uses single graphic symbols or sequences of these symbols to represent phonemes.

While the word and its meaning are conveyed intact by single signs in the logographic and the ideographic systems of writing, the syllabary and the alphabet system involve the process of putting together the various graphic symbols which form the components of a word in various levels of the hierarchy: the phoneme, the syllable, the morpheme, and the whole complete word. This process is conditioned by rules peculiar to each specific language and such a rule is termed *spelling*.

Languages using the syllabic system of spelling have somehow maintained a higher degree of uniformity in spelling than those using the alphabet system. The Javanese language of Java and the Arabic language are more consistent in their spelling systems, as each syllable unit in these languages is represented by a particular graphic shape. A syllable always consists of an obligatory vowel and one or more optional consonants. All the functional features of the vowels in the context of the language concerned are depicted in the graphic shape.

The spelling problem usually arises with the alphabet system. Languages adopting this system may at one time or another face some kind of chaos in varying degrees in their histories of development. This chaos arises mainly from the representations of the various phonemes (the functional speech sounds) by graphic symbols existing in a particular alphabet. The term *alphabet* refers to a set of graphic symbols that can represent the

*This is a revised version of a paper originally presented to the Conference of the International Federation of Free Teachers Union held in Kuala Lumpur on December 28-31, 1970.

separate vowel and consonant phonemes of a language. All alphabets are phonemically based in the sense that each graphic symbol represents at least one phoneme. Hence, the process of spelling in an alphabet system means the process of identifying the various graphic symbols with phonemes.

In speaking or reading, the phonemes are not recognised as isolatable items which are independent of their contexts. These phonemes are recognised as items only in the phonemic patterns that identify word-patterns in a given language. This same principle applies in the recognition of the graphic symbols which represent the phonemes. These symbols are responded to only in their functions as elements in graphemic patterns that visually identify the words in a given language.

The most ideal spelling system is of course one which consistently indicates a one-to-one correspondence between the phoneme and the graphic symbol or the graph such that the symbol 'b' indicates only the phoneme /b/, 'd' only the phoneme /d/, 'a' only the phoneme /a/, 'i' only the phoneme /i/, and so on. Such a spelling does not allow the representation of one phoneme by two or more symbols or two or more phonemes by one symbol. But this one-to-one phoneme-grapheme rule has never been followed with much loyalty by the natural languages of the world. This is due to the fact that most of the languages using the alphabet system are those which have adopted the Greco-Roman alphabet. This alphabet, which was devised to suit the phonemic system of the Indo-European languages, does not satisfy the needs of the other languages in representing their phonemic systems. Even in most of the Indo-European languages themselves, there are certain phonemes which do not correspond to particular graphic symbols. Hence a solution was found either by duplicating the function of one graphic symbol or by combining two symbols to denote one phoneme.

Another factor which explains the inconsistency in the spelling systems of most languages is that the spelling systems are not wholly phonemic. This is due to the fact that in a particular alphabet system, for instance, some symbols represent phonemes while others represent phones.

A phone or speech-sound is any sound which is produced by the human vocal organs, but a phoneme is a bundle of speech-sounds that are phonetically realized differently in different phonological environments. The phoneme is then an abstract concept which is physically realized as various phones.

The English alphabet is phonemically based, because the symbols or letters of the alphabet are supposed to represent some phoneme or other. On the other hand, the English alphabet is not a phonemic alphabet in the sense that there is no one-to-one correspondence between the phoneme and the graphic symbol. The English spelling system has often come under severe criticism not merely because it is not a phonemic spelling system but also because its degree of inconsistency is rather high compared to the spelling systems of other languages.

The vowel phoneme /i/ in English is graphically represented in at least 11 different ways.

- (1) as "e" — *we, me*
- (2) as "ey" — *key*
- (3) as "ee" — *bee, speed, feel*
- (4) as "e...e" — *mete, cede*

- | | | |
|----------------|---|-----------------------|
| (5) as "ea" | — | meat, lead, pea |
| (6) as "ie" | — | field, yield, believe |
| (7) as "ei" | — | seize, receive |
| (8) as "i...e" | — | machine, caprice |
| (9) as "eo" | — | people |
| (10) as "oe" | — | amoeba |
| (11) as "ae" | — | Caesar |

The combinations of sounds *ough* represent a variety of sounds and combinations of sounds, as given below:

- | | | |
|------|---|-------------------------|
| [o] | — | dough, (al)though |
| [ɔ] | — | bought, thought, caught |
| [u:] | — | through |
| [aʊ] | — | bough, plough |
| [ɔf] | — | cough |
| [ʌf] | — | tough, rough, enough |
| [ʌp] | — | hiccough |

The phoneme /f/ is represented by the symbol "f" and the sequence "ph", e.g. *father*, *fat*, *philosophy*.

The phoneme /s/ is represented by the symbols "s" and "c".

- | | | |
|-----|---|----------------|
| "s" | — | send, some |
| "c" | — | receive, cease |

The letter "c", besides representing the phoneme /s/, also stands for the phoneme /k/: *can*, *count*, *call*, etc.

One finds that in the English spelling system there are certain graphic symbols in the spelling of a word which do not represent any phoneme or speech sound as seen in the combination "ough", e.g. *though*. The syllable peak in *though* as pronounced in Queen's English is a sequence of vowels [ou] and this is symbolised by "ou". But *gh* in the spelled word does not represent any vocal symbol. In the word *rough*, the sequence *gh* can be taken to represent /f/ while in *hiccough* they represent /p/. The different phonetic realizations of *gh* in such words can only be explained by looking into their various etymologies. The same can be said of *know*, *knowledge*, *psychology*, and *gnome*, where the initial letters do not stand for any phoneme at all.

The International Phonetic Alphabet provides for this idea of having one symbol for one phoneme, but adopting this alphabet does not solve the problem as it involves new symbols and diacritics which are absent from most of the typewriters and printing machines throughout the world.

A person learning to read in English is involved in two processes:

- (1) Identifying certain graphic symbols or sequences of graphic symbols with certain phonemes, e.g. "f" is identified with the phoneme /f/ but at the same time the sequence "ph" is also identified with the very same phoneme. The sequence "oo" can be identified with either the phoneme /u/ which is characterized by the feature *length* as in *ooze*, or the phoneme /ʌ/ as in *blood*.
- (2) Memorizing the whole visual representation of a word or a syllable, e.g. *thought*, *caught*, *court*, *necessary*, etc.

On the other hand, a person learning to read in Malay has only got to master the first process. On this ground, much more time is spent on learning to read in English than in Malay. The spelling system of Malay is much more phonemic than that of English and the degree of phonemic consistency in the Malay system of spelling is much higher than that of the English system.

There are 6 vowel phonemes in Malay, namely /a/, /i/, /e/, /u/, /o/, and /ə/, and these are represented by 5 graphic symbols: "a", "i", "e", "u", and "o". Four of these graphic symbols, "a", "i", "u", and "o", indicate a one-to-one correspondence with the phonemes /a/, /i/, /u/, and /o/. The graphic symbol "e" represents two phonemes: /e/ and /ə/. This is the only inconsistency in the representation of the vowel phonemes of Malay by graphic symbols.

The same degree of consistency is found in the representation of the consonants, i.e. one consonant phoneme is represented by one grapheme. The grapheme is a graphic unit distinct from any other graphic unit. The grapheme can consist of a single graphic symbol, such as "a", "n", "g", etc., which are distinct from each other; or it can also consist of a sequence of graphic symbols, e.g. "ny", "ng", "ch", and "sh", which contrast with other sequences of graphic symbols. The sequence "ny", "ng", "ch", and "sh" stand for the phonemes /ɲ/, /ŋ/, /tʃ/, and /ʃ/. The component of each of the graphemes "ny", "ng", and "sh", when taken separately, already has a function in the representation of a particular phoneme. E.g. the symbol "n" represents the phoneme /n/, and "y" represents /j/. Thus these combinations are justified. There is no phonemic or orthographical justification in the choice of the combination "ch" for the phoneme /tʃ/. The symbol "h" stands for the phoneme /h/, but "c" by itself does not stand for anything at all in the Malay spelling system. The only reason for the retention of "ch" is its long-imposed existence in the Malay alphabet.

The early orthographers of Malay could have given more credit to themselves if they had chosen just the symbol "c". The obvious explanation for the choice of "ch" is the wholesale transfer from the English way of representing their phoneme /tʃ/, as in *change*, *charge*, *much*, but then the English spelling system has its own reason for using "ch" because "c" by itself in that system has already symbolized other phonemes – /k/, as in *cold* and *can*, and /s/ as in *receive* and *centre*.

Loanwords from Arabic have brought with them three more graphemes of the same nature as the ones discussed above:

(i) "dh", which stands for the Arabic ض, the pharyngealized [d̤] as in *hadhiri*, "to be present at", *fardhu*, "compulsory deed".

(ii) "th", which stands for the voiceless dental fricative [θ], as in [θəlaθa].

(iii) "dz", which stands for the voiced dental fricative [ð], as in *bedza*.

The phones [d̤], [θ], and [ð] are not phonemes in Malay, and as such they are realized in the speech of Malay speakers as [d], [s], and [z], respectively. On this basis, the graphemes "dz" and "th" have been fully replaced by "z" and "s".

However inconsistent the English system of spelling is, as far as the graphological representation is concerned, it does have some uniformity. This means that although the spelling of a word may not be plausible in its linguistic as well as in its orthographical aspects, this word is spelled only in that way, and its orthographical representation is not duplicated. Thus the word *necessary* has only one spelling as are the vast majority of the

other words. Hence one can say that the English system of spelling is consistent in its inconsistency.

In the past few decades, some changes have taken place in the English system of spelling stemming from America. Hence *coconut* is no longer spelled with an "a" before "n" – "cocoanut". *Program* has dropped off the last two letters, *me*; the final *-ise* in a verbal form such as *characterise*, *formalise*, etc., have given way to *-ize*, and *ae* in *aesthetics* has been merged into "e".

Although the Malay language can boast of a high degree of phonemicity in its spelling systems, it cannot claim having the same degree of uniformity. There are at least four spelling systems now used in writing the Malay language. The one used throughout the schools of Malaysia is the *Ejaan Sekolah*, also known as Ejaan Zaaba, but outside the school compound much freedom is given to individuals to choose any of the systems in current usage, and these are Ejaan Sekolah, Ejaan Wilkinson, Ejaan Kongres, and Ejaan Bersama Malaysia-Indonesia. These systems differ from each other in the following aspects:

- (i) The choice of symbols to represent the vowel phonemes /i/ and /u/ in closed final syllables of words.
- (ii) The use of the hyphen.
- (iii) The writing of the clitic forms and the particles.
- (iv) The writing of reduplicated forms.

Let us briefly examine these aspects one by one:

(1) *The vowel symbols*

The vowels /i/ and /u/ in closed final syllables of words have realizations which differ slightly from their realizations in open syllables. In open syllables, these vowels are really closed vowels, e.g. /titi/, *titi*. In closed syllables their realizations indicate some degree of openness, i.e. [i] and [i<] as in [bili<ʔ]. Thus both the highly closed [i] and the open [i<] are allophones of a single phoneme /i/. The same goes for the phoneme /u/. Hence only one symbol is needed for each of these two phonemes, /i/ and /u/, no matter where it stands in a word pattern.

In the Ejaan Sekolah the slightly open [i] in closed syllables is taken to be a member of the /e/ phoneme, as its phonetic realization is almost similar to one of the realizations of /e/, while the slightly open [u] is taken to be a member of the /o/ phoneme. This conception is manifested in the rule laid by the system which provides for the determination of "e" and "o" in closed syllables. The rule says that the symbol has to be either "h" or "k" before "e" and "h", "k", "r" or "ng" before "o". In all other cases of closed syllables, the vowel symbol should be "i" or "u". We see here that the consonantal environments which are supposed to determine the graphic occurrence of "i" and "u" are not linguistically based. The symbols "h", "k", and "ng" are supposed to represent velar phonemes /h/, /k/, and /ŋ/, but the symbol "r" represents the alveolar trill /r/. The phonetic value of /r/ is closer to that of /l/, /n/, and /s/ (all being alveolar consonants) than to the velar consonants /h/, /k/, and /ŋ/. Why then has /r/ got to be treated differently from its alveolar counterparts? In this aspect, Ejaan Wilkinson, Ejaan Kongres, and Ejaan Bersama Malaysia-Indonesia* are phonemically more plausible in the sense that they consistently

*Ejaan Bersama Malaysia-Indonesia is the spelling system agreed upon by Malaysia and Indonesia in 1967. The system agreed upon by the two countries in 1959, better known as the Malindo System, has never been made public.

use the symbols “i” and “u” to represent the phonemes /i/ and /u/ in all environments no matter what their phonetic realizations are. The choice between “i” and “e” on the one hand and “o” and “u” on the other is the only parting line between the Ejaan Sekolah and the Ejaan Wilkinson.

(2) *The Use of the Hyphen*

The Ejaan Sekolah and Ejaan Wilkinson make more use of the hyphen than the two other systems. The use of the hyphen in some cases does not have any structural or orthographical basis. By orthographical basis I mean a basis which pertains to a visual representation which facilitates the reading of the spelled word.

Certain prefixes are separated from their root-words by hyphens but others are not. The verbal prefix *di-*, which occurs in passive verbs, is hyphenated from the root-word, obviously because it is homophonous with the locative particle *di*, and this particle is of course separated from the word that follows it by a hyphen. The prefix which is written in the same way as *di-* is *se-*, but *se-* is not homophonous with any particle. It is my conjecture that the formulators of the systems placed the Malay prefix *se-* in the same mould as the English article *a*, as in the phrase *a book*. But then the meaning “one” conveyed by “*se-*” is only confined to the nominal *se-*. The adjectival *se-* and adverbial *se-* convey other meanings. And linguisticians will always point out that meaning should not be taken as the primary criterion in linguistic analysis.

Other prefixes are never hyphenated from their root-words. This differential treatment employed in the graphic representation of the prefixes has caused a lot of confusion in the teaching of Malay grammar. Ejaan Kongres and Ejaan Bersama Malaysia-Indonesia have done away with the hyphen. In these systems, the prefixes are joined to the roots.

(3) *The Writing of the Clitic Forms and the Particles*

The clitic forms are *ku*, *kau*, *mu*, and *nya*. The first two can enter into a construction with a preceding or a following word while the latter two can only enter into a construction with a following word.

The particles which turn out to be problematic in the spelling system are *di*, *ke*, *lah*, *tah*, and *pun* which are constructed only with preceding words. In their relationship with the words with which they form constructions, the particles as well as the clitic forms are represented in various ways:

- (a) They are separated by a hyphen from the words concerned.
- (b) They are written as single words in their own right.
- (c) They are joined to the words concerned.

(4) *Reduplicated Forms*

The reduplicated forms in the Malay spelling systems are represented in two ways:

- (a) By placing the numerical symbol “2” after the form to be reduplicated.
- (b) By writing the reduplicated form in toto.

The lack of uniformity in the Malay spelling system has created some confusion in the teaching and writing of the Malay language and there has always been an atmosphere of liberalism in this field. With this situation, a prescriptive standpoint is difficult to achieve.

What the Malay language needs at the moment is one and only one system of spelling,

a system which should be implemented at all levels of teaching as well in all fields of written Malay. This system, besides being uniform, should have the following characteristics:

- (i) As far as possible, it should be a phonemic spelling.
- (ii) It should be practical in the sense that it does not introduce new symbols or new graphemes and that diacritics are used to the barest minimum.
- (iii) It should not contain any representation which is grammatically or orthographically inexplicable.
- (iv) It should be flexible in the sense that it provides for the inclusion of loan words which contain loan phonemes and innovations in the phonological structure, such as certain clusters which, prior to the introduction of words borrowed from the Indo-European languages, were alien to the Malay phonological as well as the spelling systems. Such clusters occur in words like *proses*, *stereo*, *sains*, etc.

A compromise between the existing systems and an improvement on the compromised form will result in a system which has all the four characteristics mentioned above. And working on these four characteristics, the University of Malaya in Kuala Lumpur has since January 1971 come out with its own spelling system known as *Ejaan Universiti Malaya 1971*. This system was formulated on the basis of the *Ejaan Bersama Malaysia-Indonesia* and the *Ejaan Chadangan*. The *Ejaan Chadangan* is a modified version of the *Ejaan Bersama Malaysia-Indonesia*, and the person responsible for this modification is Mr. Hassan Ahmad, Director of the Language and Literary Agency in Kuala Lumpur.

The University of Malaya Spelling System bears a great resemblance to the *Ejaan Chadangan*. They differ in the following:

- (i) The spelling of certain words.
- (ii) The presence of rules in the spelling of loan words (particularly the technical terms) in the University of Malaya Spelling System compared to the absence of such rules in *Ejaan Chadangan*.

It can also be said that the University of Malaya Spelling System is close to the Spelling System of Bahasa Indonesia, differing only in:

- (i) the use of certain graphemes, and they are as follows:

<i>University of Malaya System</i>	<i>Bahasa Indonesia System</i>
y	j
j	dj
sh	sj
ch	tj
kh	ch

- (ii) the spelling of certain words.

The formulation of the University of Malaya Spelling System was part of the implementation program of the National Language Policy which was launched in November 1971. The implementation of this system of spelling in the University of Malaya campus has brought about an atmosphere of confidence among the administrators, secretaries, clerks and typists who are very much involved in preparing correspondence, reports, and minutes of meetings. The system also proves useful to the various terminology committees of the University, specifically in the spelling of loan scientific terms.

At the beginning the public reaction toward the University of Malaya Spelling System

was one of hostility. Certain newspaper editorials emitted severe criticisms against those responsible for the formulation of the system. The public, confused by the many systems already in existence, was of the opinion that another system would aggravate the situation. This was the feeling borne by the public before they ever set eyes on the system. Once they did, however, they were drawn toward it and have come to adopt it on their own free will. Further support for this system was seen in the meeting between publishers, writers, and representatives of institutions of higher learning and mass-media held in Kuala Lumpur in the premises of the buildings of a leading Malay newspaper group, the Utusan Melayu Press, in the third week of January 1971. In the Congress of Malay Teachers held in Kuala Lumpur in April 1971, a resolution was adopted to urge the government to enforce the implementation of a single spelling system for the Malay language, and in search of this system, which was to be the standard system, the University of Malaya Spelling System was to be taken as the basis of reference.

The National Cultural Congress held in Kuala Lumpur in August 1971 also passed a resolution to urge the government to expedite the implementation of a standard spelling system. To date, the Malaysian public is still waiting for the government decree to that effect.

Appendix

The following are my own publications which to some extent discuss the spelling problems in Malaysia, the efforts to standardize the systems in Malaysia and Indonesia, and the historical backgrounds of the various systems:

1. "Towards the Unification of Bahasa Melayu and Bahasa Indonesia: An account of efforts to standardize the spelling systems of Malay in Malaysia and Indonesia," *Tenggara*, No. 1, University of Malaya, Kuala Lumpur.
2. "Standard Language and the Standardization of Malay," *Anthropological Linguistics*, February 1971; *Suara Gabongan*, Vol. 1. No. 6, 1971, pp. 29–42.
3. "Bahasa Standard dan Standardisasi Bahasa Melayu," *Dewan Bahasa*, Jilid XIV, Bilangan 10, 1970, hal. 435–455.
4. "Perkembangan Bahasa Malaysia Dan Ejaan Baru," *Dewan Bahasa*, Jilid XV, Bilangan 5, 1971, hal. 195–204.
5. "Penyelarasan Sistem Ejaan Bahasa Malaysia Dalam Konteks Perkembangan Bahasa Malaysia," *Guru* Jilid IV, Bilangan 2, June 1971, hal. 45–54.