

# Projects and Activities on Lexicography in Dewan Bahasa Dan Pustaka: Past, Present and Future

Noresah Baharom

*Malaysia*

## 1. Introduction

This paper aims to present a brief scenario of projects and activities pertaining to lexicography in which Dewan Bahasa dan Pustaka (DBP) has been, is currently, and will be actively involved in the near future. However a major part of the discussion in this paper will be focussed on current and future projects, with special reference to activities of updating the existing authoritative dictionary of the Malay language (the *Kamus Dewan*) and those in connection with corpus and structured text file (lexical data base) creations to facilitate the compilation of a comprehensive dictionary of the Malay language, *Kamus Besar Bahasa Melayu*. Issues and problems encountered in developing the database will be among other things highlighted. A brief description of the latest dictionary publication of DBP, a comprehensive English-Malay bilingual dictionary, deserves a mention too in this paper. Before going in-depth on projects, a brief introduction on the roles and objectives of DBP will be given.

## 2. The Roles Of Dewan Bahasa Dan Pustaka (DBP)

Dewan Bahasa dan Pustaka was established in 1956 as a government bureau in the Education Department, and later transformed into a statutory body vested with the authority to carry out the following functions:

- 1) To develop and enrich the Malay language, which is the national and official language of Malaysia, and also the language which is used as the medium of instructions in schools and universities in Malaysia. The Malay language also plays the important role of unifying the multiracial people of Malaysia.
- 2) To promote creativity and creative talents.
- 3) To publish books of various kinds and invarious subject fields in the national language.

Among the roles entrusted to DBP in order to develop and enrich the Malay language are language planning, which includes corpus planning and corpus development, and language codification and cultivation. Besides the creation, standardisation, and dissemination of terminologies of various disciplines, and of the orthographic system, DBP is also entrusted to codify the language by producing various kinds of dictionaries, namely monolingual, bilingual, and technical dictionaries.

## 3. Lexicography Projects In DBP

### 3.1 Introductory Statement

As stated above, DBP is actively involved in three basic programmes of lexicography work, the monolingual, bilingual, and technical programmes. In the early days of its establishment, the Lexicography Division of DBP focused on monolingual and bilingual lexicography. As such the

first dictionary that was produced by DBP was the *Kamus Dewan* which, until now, is still regarded as the most authoritative monolingual dictionary of the Malay language. The second dictionary that was published by DBP was the *Kamus Dwibahasa* which is an English-Malay dictionary. The *Kamus Dewan* was published in 1970, and the *Kamus Dwibahasa* was published in 1979. Work on technical dictionaries as set aside for quite some time until work on the creation of terminologies of various disciplines reached a certain level of standardisation; work on the standardisation of terminologies has been going on since December 1972 in collaboration with the Indonesian government, and later in 1985 with the Brunei Darussalam government as well under the name of MABBIM (an acronym which stands for Majlis Bahasa Brunei Darussalam-Indonesia-Malaysia). The publication of *Kamus Istilah Sains Am* marked the publication of the first technical dictionary by DBP. In the years that followed, more technical dictionaries of various fields of study emerged, and this activity of increasing the number of technical dictionaries is still going on indefinitely. This year, 1992, marked the birth of another bilingual English-Malay dictionary by DBP, a dictionary which took 13 years of laborious painstaking editorial work. This dictionary consists of more than 38,000 main entries. A list of dictionaries so far published by DBP is given in the appendix of this paper.

### 3.2 Revising the *Kamus Dewan*

The *Kamus Dewan* which was first published in 1970 using the old spelling system of the Malay language was first reprinted using the new spelling system in 1984, and later revised to produce its second edition in 1989, with an additional 2,000 lemma (its first edition consists of approximately 28,000 lemma). Due to the rapid development of the Malay language as the medium of instruction for all levels of education, and also as the language of unification of the multiracial society of Malaysia, there has been an urgent need to record lexical changes that have now become a salient feature of the language, in the shortest time possible. Two alternatives were carefully considered. The first is to revise the existing dictionary, and the second is to write a totally different and new dictionary. Since the second alternative would take a much longer time, the first alternative was given priority, while work on preparing the concept and foundation of a more comprehensive dictionary of the language is simultaneously handled.

Basically the updating of *Kamus Dewan* is focussed on two major aspects: the addition of new words or words with new usage, and the revising of existing data, particularly those in connection with accuracy of facts and of grammatical information, minor changes in spelling, inconsistent treatment of homonyms and of the dictionary format in general.

As in any other language, lexical change in the Malay language occurs due to the addition of new words, changes in the meanings of words, and the loss of words. New words enter the language by compounding (the recombining of old words to form new ones with new meanings), such as **pemegang amanah** 'trustee' and **pemegang saham** 'shareholder'; by derivational processes, such as **pensejagatan** 'universalisation', which is a derivative formed from the base word **sejagat**; by word coinage, i.e., deriving words from acronyms, such as **ikwam** (Institut Kemajuan Wanita Malaysia) and from abbreviations, such as **ISBN** (International Standard Book Number); and also by borrowing. With the very rapid flow of information technology, millions of loan words get into the Malay vocabulary, and this of course reflects the necessity to respond to the information explosion that is experienced everywhere else in the world in the fastest way possible. Generally, in the current situations, most loan words in Malay are borrowed from the English language, and some are taken from the Arabic language, particularly words relating to religious matters. Loan words are generally spelled and pronounced according to the Malay spelling and phonemic systems. Among new words that are collected for the third edition of *Kamus Dewan* are **genesis** 'genesis', **keheterogenan** 'heterogeneity', and **beridentitikan** 'having

something as an identity'. Slang words created, especially those used by teenagers, are also considered as new entries for the new edition. The collection of dialectal words too forms part of the activity to accumulate new words for the new edition of the dictionary. Among existing words which have acquired new or additional meanings in the course of time which are also treated in the new edition are **wawasan** currently and popularly used as conveying the concept of 'vision' instead of 'insight' as generally understood a few years ago, and **canggih** which originally meant 'talking too much', but now has expanded its meaning to include the concept of 'sophistication'. The later meaning carried by this word was at one time associated/collocated with the word technology only. However recent development of the language shows that because of the general acceptance and popularity of the word, it now seems to collocate with almost everything—it seems to be readily collocated with nouns like restaurant, dressing, atmosphere, furnishing, fashion, etc.

One obvious difficulty encountered in revising this dictionary is the decision to treat obsolete words and phrases/sentences. The dictionary was originally compiled using corpus data that were published before 1960s, while the revising activity is basically focussed on current usage of the language. As such it is inevitably impossible not to discover words which are not actively used anymore in the existing dictionary—words which are hardly recognised by the younger generations. It was finally decided not to amend any of these words for historical reasons, and it is in relation to this situation that the idea to write a totally new dictionary was proposed—a more comprehensive monolingual Malay dictionary, the *Kamus Besar Bahasa Melayu*, mentioned earlier.

### 3.3 The Corpus and Structured Text File Creation

As mentioned above, there has been a growing need to produce a more comprehensive dictionary of the Malay language because of the rapid development of the Malay lexicon, both in terms of size and semantics (generally semantic expansion), and also of the obsolete nature of some of the word forms and phrases/sentences in the existing dictionary. In line with the modern approach of dictionary compilation, DBP is currently embarking on a dictionary project using a computer corpus as a collection of source material for the dictionary. Various criteria are considered in designing the corpus, such as corpus size, sources of materials used (this includes books, journals, magazines, brochures, newspapers), language mode (written versus spoken), genre, and sex of authors. To date, the size of the developing Malay corpus is about 7 million words. It is targeted to reach about 10 millions by the end of 1992, before the analysis of the generated concordance is scheduled to commence. While work on corpus creation is going on, work on developing the corpus system to retrieve and manipulate the corpus data is simultaneously carried out. A prototype corpus system has been successfully developed with the assistance of computer experts from University Sains Malaysia. The prototype system is able to display several attributes of a particular text file under study, such as the source title, name of author, year of publication, type of text (original writing or translation), genre etc. Besides these attributes, the system is also able to generate concordances of a particular word form, which is the key tool to lexicographers. However there has been several limitations observed about the system; it is not able to accommodate large corpus data since it is developed on an ordinary PC, and that the lines of concordance retrieved for a particular word form do not actually represent the overall corpus data; instead it represents data from a particular text file or a few text files. These limitations are expected to be overcome with the availability of work stations in the near future.

Work on structured text file creation is expected to commence in late 1993, when the system is ready to fulfil the various criteria outlined for lexicographical needs.

### **3.4 Dewan English-Malay Dictionary**

#### **3.4.1 Historical Information**

The initial impetus for compiling this dictionary came from a British philanthropist, Tan Sri John Galvin, who provided most of the funds during the early years of the project. The actual compilation was begun in 1955 by Dr. W. Linehan. The project was carried on by C.C Brown, a lecturer in Malay from the University of London. Brown submitted the manuscript to Cambridge University. Later it was sent to Professor A.H. Johns of the Australian National University to be edited and revised.

In 1967, a team headed by Professor A.H. Johns started compiling a new English-Malay dictionary but work on the project had to stop due to lack of funds. A year later, DBP officially took over the final manuscript, and the compilation work was completed in January 1992.

#### **3.3.2 Principal Features of the Dictionary**

##### **Meaning Discrimination**

The dictionary is intended for English speaking users, particularly translators. Hence, a great deal of attention is given to meaning discrimination, especially of polysemous words to help the reader find the appropriate equivalent, distinguish one meaning or homograph from the other and use the word he is looking up in the correct way. The meaning discriminations (or glosses) are very brief, and placed before the Malay equivalents. These glosses give the reader extra information on the entry; for instance information on collocations aids the reader to make the correct choice of equivalent for a particular context.

##### **The Equivalent**

Since the dictionary is basically intended for translators, a great deal of effort had been made to ensure the correctness and appropriateness of the Malay equivalents provided. The dictionary consciously avoids heaping up a whole range of equivalents. If there is a choice, explicit directions are given on how to select the right equivalent. If there is no equivalent for a particular word, the entry may be omitted, or a paraphrase of the meaning is given or advice is given on the use of the untranslated English word.

##### **Example Sentences**

There are more than 50,000 example sentences, selected from about 150,000 English sentences, which are the bedrock of this dictionary. The sentences try to show typical usage. They are intended to help users in translation and sentence production.

##### **Word Class**

Every headword in the dictionary is given a word class. The word class is written after the headword, variant spelling or a derivative of the headword. In an entry that has more than one word class, the different word classes are dealt with in alphabetical order.

Where necessary, usage labels (colloquial, taboo, formal, archaic, rare, euphemistic), regional labels (US, New Zealand, etc.) and subject labels are used to mark a whole entry or an individual phrase or meaning.

##### **Technical Terms**

Since specialised vocabulary terms are continually creeping into our conversation, reading and writing, many of these regularly used terms are included in the dictionary.

#### 4. Conclusion

The above description shows that DBP, as the literature and language agency of Malaysia, has been and will be continually involved in compiling various kinds of dictionaries for the codification and development of the Malay language, the official and national language of Malaysia. With the sophistication of computer technology, the effort to create corpus material and structured text files for bilingual dictionaries using the existing comprehensive English-Malay dictionary data (which is already in the electronic form) will no doubt become a reality in the near future.

## APPENDIX

## Dictionaries Published by Dewan Bahasa dan Pustaka

NO.	TITLE	YEAR OF PUBLICATION	YEAR OF REPRINT	NEW EDITION
1.	Kamus Dewan (ekabahasa)	1970	1984	1989 (currently being revised for the 3rd. edition)
2.	Kamus Dwibahasa	1979	1981,1985	(currently being revised for the 2nd edition)
3.	Kamus Bahasa Malaysia (Edisi Pelajar)	1975	1983	
4.	Kamus Pelajar	1987	1987,1988,1989, 1990,1992	
5.	Kamus Istilah Sains Am	1988		
6.	Kamus Rendah Dewan Bergambar	1989		
7.	Kamus Fizik: Mekanik	1989		
8.	Kamus Kimia: Biokimia	1989		
9.	Kamus Biologi: Mikrobiologi	1990		
10.	Kamus Biologi: Hitologi dan Sitologi	1990		
11.	Kamus Biologi: Fisiologi	1990		
12.	Kamus Fizik: Elektromagnet, Keelektrikan dan Kemagnetan	1990		
13.	Kamus Fizik: Optik	1990		
14.	Kamus Matematik	1990		
15.	Kamus Biologi: KBSM	1990		
16.	Kamus Perhutanan	1990		
17.	Kamus Kimia: KBSM	1991		
18.	Kamus Mekanik Gunaan	1991		
19.	Kamus Perakaunan	1991		
20.	Kamus Perubatan	1991		
21.	Kamus Inggeris-Melayu Dewan	1991		
22.	Kamus Biologi: Patologi	1991		
23.	Kamus Kimia: Kimia Analisis			
24.	Kamus Kimia: Kimia Fizik dan Kimia Polimer			

REFERENCES

- Asmah Hj. Omar. 1975. *Essays on Malaysia Linguistics*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- . 1975. *Perancangan Bahasa dengan Rujukan khusus kepada Perancangan Bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Carter, R.A. 1987. *Vocabulary: Applied Linguistic Perspectives*. Allen & Unwin Ltd.
- Cobuild. 1987. *Collins Cobuild English Language Dictionary*. William Collins Sons & Co. Ltd.
- Dewan Bahasa dan Pustaka. 1989. *Kamus Dewan*.
- . 1991. *Kamus Inggeris Melayu Dewan. An English Malay Dictionary*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- . 1992. Majlis Bahasa Brunei Darussalam - Indonesia - Malaysia (MABBIM). Unpublished.
- Fromkin, V. and R. Rodman. 1988. *An Introduction to Language*. 4th. Edition. Saunders College Publishing.
- Hassan Ahmad. 1988. *Bahasan Sastera Baku: Cetusan Fikiran*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Krishnamurthy, R. 1987. *Looking up: An Account of the Cobuild Project in Lexical Computing*. Edited by J. M. Sinclair. William Collins Sons & Co. Ltd.
- Noresah. n.d. "Corpus Evidence in Malay Lexicography: A feasibility Study." M.A. thesis. Unpublished.
- . 1992. "Perkamusan Melayu dalam Memenuhi Wawasan 2020." Paper presented in the International Conference on Malay Linguistics. Kuala Lumpur: (11-13 August 1992).
- Palmer, F.R. 1981. *Semantics*. 2nd edition. Cambridge University Press.
- Sinclair, J.M. 1987. "The Nature of the Evidence." *Looking Up: An Account of the Cobuild Project in Lexical Computing*. Williams Collins Sons & Co. Ltd.