Malay Framenet: An Application for Bilingual Dictionary

Intan Safinaz Zainudin

School of Language Studies and Linguistics, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ABSTRACT

Boas (2009) has established that the recent trends in computational lexicography encompass computational methods and tools designed to assist in various lexicographical tasks, including the preparation of lexicographical evidence from many sources, the recording in database form of the relevant linguistic information, the editing of lexicographical entries and the dissemination of lexicographical products. In line with this development, Intan Safinaz’s research (2011) on frame-based analysis on Malay corpus has produced essential semantic information compiled in a reference database that can assist lexicographers in editing dictionary entries for Malay-English bilingual dictionaries. This paper aims to discuss the frame-based analysis that incorporated Frame Semantics methodology and the benefits of compiling the database as a reference database. The entry ‘memberi’ is analysed and the findings contribute to the semantic content of the reference database. This reference database is able to facilitate the task of editing entries of the Malay monolingual dictionaries, Malay-English bilingual dictionaries, as well as the Malay language reference books. This paper also discusses the possibility of building a Malay FrameNet based on the architecture of the English FrameNet which was set up at the International Computer Science Institute in Berkeley. The proposed methods of building the Malay FrameNet is also inspired from the FrameNets of other languages for example the German, French and Japanese FrameNets. The setting up of the Malay FrameNet can be seen as an effort towards processing the Malay language and strengthening the Malay computational lexicography as well as encouraging future linguistic and computational lexicography research and cooperative efforts in natural language processing of the Malay language.

Keywords: Computational lexicography, bilingual dictionary, Frame Semantics, FrameNet, frame-based analysis
INTRODUCTION

Dictionary research on bilingual Malay-English dictionaries has not been given due attention. Although some researchers from the Dewan Bahasa dan Pustaka or Malaysian Institute of Language and Literature have paid attention to the lack or non-existence of Malay equivalents for English lexicals in English-Malay bilingual dictionaries (Nor Azizah & Azizah 1993, Noresah & Fadilah, 2003; Fadilah, 2005), there is very little research done on the setting up of a database for dictionary building in the local scene. An attempt was first explored by Lim Tian Tze, Tan Ewe Hoe and Tang Enya Kang (2006) in their effort to digitise the monolingual Malay dictionary Kamus Dewan based on Wordnet lexical database. The Malay WordNet database was proposed to compile Malay lexical resources to facilitate more meaningful operations that could also be aligned with other languages. However, Lim et al.’s research is limited to building a database based on a monolingual dictionary. This transformation and innovation in dictionary research must be extended to research on bilingual dictionary. Therefore, this paper attempts to describe a proposal of a Malay database for the purpose of building up or editing a bilingual dictionary, specifically Malay-English dictionary.

Berkeley’s FrameNet, an English language lexical database, was chosen as the template database to build a Malay FrameNet because of its suitability in processing Malay data. Furthermore, Berkeley FrameNet has been used as a template for other world languages like the SALSA project or German FrameNet (Burchardt et al., 2009), Japanese FrameNet (Ohara 2009), and others that will be discussed briefly. The Malay FrameNet is proposed because it can accommodate Malay semantic information as a reference database that can be aligned to the English Berkeley FrameNet. The proposed database contains bilingual information that can assist lexicographers in building as well as editing dictionary entries for Malay-English bilingual dictionaries.

This present paper focuses on the importance of the setting up of the Malay FrameNet as a reference database that involves other ‘computational methods’ (Boas 2009) such as the editing of lexicographical entries, the recording in database form of relevant linguistic information and the dissemination of lexicographical products, among others. This paper begins with a brief discussion on the FrameNet, the theory of Frame Semantics and other non-English FrameNets. Then, it proceeds with a discussion on how a frame-based analysis was carried out on 5 million word Dewan Bahasa dan Pustaka and Universiti Kebangsaan Malaysia (DBP-UKM) Malay corpus and how the Frame Semantics methodology was adopted. The frame-based analysis also showed how the Malay transitive verb ‘memberi’ evoked five Malay frames and how the Malay frames proposed a revised meaning for the dictionary entry ‘memberi’. The frame-based analysis also suggested the English equivalents for ‘memberi’ after mapping the
Malay and the equivalent English frames from the Berkeley FrameNet. The findings of the analysis consequently contribute to the semantic information for the proposed database. A short discussion on matters pertaining to the structure of the Malay FrameNet will follow at the end of this paper.

**REVIEW OF LITERATURE**

This section reviews the related literature on FrameNet and other non-English FrameNets.

*What is FrameNet?*

FrameNet is an online English language lexical database. FrameNet can be accessed at [https://framenet.icsi.berkeley.edu](https://framenet.icsi.berkeley.edu) and it was built based on the principles of Frame Semantics (Baker *et al*., 1998; Baker *et al*., 2003). The FrameNet project was set up at the International Computer Science Institute in Berkeley, California, and this database describes several thousand words of English language that are taken from a variety of semantic domains from the British National Corpus and other corpora. FrameNet identifies and describes semantic frames and analyses the meanings of words by appealing directly to the frames that underlie their meanings. Isolated corpus sentences are utilised to support the meaning of each frame and its frame elements.

The FrameNet project is also known as the Berkeley FrameNet, and built on Frame Semantics, a theory developed by Charles Fillmore and his associates over the past three decades. It builds on a common background of the knowledge or semantic frames against which the meaning of words is interpreted. Fillmore and Atkins's study (1992) on ‘risk’ started off with a clear definition of the central concepts underlying Frame Semantics.

According to Boas (2009), FrameNet differs from other approaches to lexical description such as WordNet because frames make use of independent organisational units that are larger than words, i.e., semantic frames (see Atkins, 2002; Boas, 2005; Atkins & Rundell, 2008). As such, FrameNet facilitates a comparison of the comprehensive lexical descriptions and their manually annotated corpus-based example sentences with those of other lexical units or LUs belonging to the same frame. Another advantage of the FrameNet architecture lies in the way the lexical descriptions are related to each other. The use of detailed semantic frames, which capture the full background knowledge evoked by all LUs of the same frame, makes it possible to systematically compare and contrast the numerous syntactic valence patterns (see Atkins, 2002; Boas, 2005, 2009).

**Reviews on Other Non-English FrameNets**

After the setting up of the FrameNet in Berkeley, studies on Frame Semantics and FrameNet have gained popularity among linguistic scholars and lexicographers around the world, hence the building of many non-English FrameNets. Boas (2009) describes a few important aspects that need to be considered when building a FrameNet:
1. The text coverage – the coverage of corpus text that needs to be annotated, whether the entire corpus needs to be semantic annotated or just isolated corpus sentences are identified to support the meaning of frames;

2. The types of sources as data pools, whether to use existing dictionaries or databases and whether to manually annotate just like Berkeley FrameNet or automatic annotation like French FrameNet;

3. A particular or general semantic domain - As in the case of Kictionary, Shmidt (2009) focuses on lexical domains of football language. However, a majority of the non-English FrameNets aim to create a database with a broad coverage;

4. The available choices of tools or methodologies; for example, the Berkeley FrameNet offers for purchase the soft copy version of FrameNet that can be translated into other languages.

Boas (2009) also summarizes the structure and development of multilingual FrameNets from around the world at centres and universities, such as:

1. The SALSA project or German FrameNet (Burchardt et al., 2009) is interested in a full-text annotation of an entire corpus. SALSA uses its own software to conduct a semi-automatic annotation on top of the existing syntactic annotation found in the TIGER (a German corpus).

2. The Japanese FrameNet uses different types of resources of data pools, apart from exploiting a monolingual corpus (Ohara, 2009). The Japanese FrameNet uses a manual annotation and adopts the Berkeley software with slight modification.

3. The French FrameNet focuses on the semantic information in multilingual databases and employs multilingual corpora (Fontenelle, 2009) and other existing lexical resources. The French FrameNet uses a semi-automatic annotation and also translates the Berkeley FrameNet into French.

4. The Spanish FrameNet uses Berkeley software with a slight modification (Subirats, 2009).

5. The Hebrew FrameNets integrates off the shelf software packages (Petrick, 2009).

FRAME SEMANTICS (FS): THE THEORETICAL FRAMEWORK

Frame Semantics or FS’s main proponent is Charles Fillmore (1982), together with other cognitive linguists, based at the University of California at Berkeley. Frame or domain can be viewed as a system of related concepts.

According to Fillmore, Johnson and Petrick (2003):
Frames are schematic representations of the conceptual structures and patterns of beliefs, practices, institutions, images that provide a foundation for meaningful interaction in a given speech community.

Frames are background knowledge as well as conceptual structures modelling prototypical situations (Burchardt et al., 2006). Table 1 below sums up the frame STATEMENT which describes a communication situation, evoked by verbs such as ‘acknowledge or admit’. Each frame has its own semantic roles called frame elements (FEs). The FEs for the frame STATEMENT are Speaker, Addressee of the statement, the Message and the Topic (Burchardt et al., 2006), as shown in Table 1 below.

TABLE 1
Frame STATEMENT

<table>
<thead>
<tr>
<th>Frame STATEMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>This frame contains verbs and nouns that communicate the act of a Speaker to address a Message to some Addressees using language. A number of the words can be used performatively such as ‘declare and insist’.</td>
</tr>
</tbody>
</table>

Frame Elements (FEs)

- **Speaker**: Evelyn said she wanted to leave.
- **Message**: Evelyn announced that she wanted to leave.
- **Addressee**: Evelyn spoke to me about her past.
- **Topic**: Evelyn’s statement is about her past.
- **Medium**: Evelyn preached to me over the phone.

Lexical Units (LUs)

- Acknowledge, v, acknowledgement, n. add. v, address. v, admission, n. admit. v, affirmation, n, allegation. n, allege. v, announce. v, announcement. n, assert. v, assertion. n, attest. v,

In Frame Semantics, the semantics is also linked to the syntactic structure of the lexical item called the theory of valency. According to Fillmore and Atkins (1992), ‘each lexical item or idiomatized phrase can be associated with what can be called as its valence description, a description that specifies, in both semantic and syntactic terms, what the expression requires of its constituents and its context, and what it contributes to the structures that contain it. The most developed systems of valence descriptions concern with the grammar and meaning of verbs’.

THE FRAME-BASED ANALYSIS OF ‘MEMBERI’ BASED ON THE MALAY CORPUS

Intan Safinaz’s research (2011) on the frame-based analysis of the Malay corpus produced essential semantic information compiled in a reference database that could assist lexicographers in editing dictionary entries for Malay-English bilingual dictionaries. The entries of ‘memberi’ from the three dictionaries were compared, as shown in Table 2.

The three dictionaries were Kamus Dewan Edisi 4 or KD4 (Malay monolingual), Intelek Malay-English Dictionary or IMED and Kamus Perwira Bahasa Melayu-Bahasa Inggeris or KPBMII (Malay-English bilingual). The comparison showed inconsistencies of the meaning coverage of the ‘memberi’ entries. In addition to that, the English equivalents for ‘memberi’ provided in the bilingual dictionaries were inadequate and unsuitable to the context of meanings.

The frame-based analysis of ‘memberi’ identified five Malay frames and the frame elements or FEs. The 5 Malay frames are:
1. GIVING PHYSICAL THEME;
2. GIVING ABSTRACT THEME, that is, the parent frame to two other frames PERMISSION and ATTENTION;
3. COMMUNICATION;
4. CAUSATION; and
5. PERMITTING

In this paper, only the frame GIVING PHYSICAL THEME is discussed. The GIVING PHYSICAL THEME frame consists of 3 core Fes, namely, Donor, Theme, and Receiver. The Donor presents a Theme of a physical entity to the Receiver. The example sentence illustrating the frame of GIVING PHYSICAL THEME below is sourced from the DBP-UKM Malay corpus:

Wakil Rakyat pula memberi hadiah kain pelikat dan kain batik kepada penduduk kampung yang miskin.

Donor – Wakil Rakyat/ Member of Parliament
Node - memberi / gives out

TABLE 2
Entry ‘memberi’ from a monolingual Malay dictionary KD4 and two Malay-English bilingual dictionaries, IMED & KPBMBI

<table>
<thead>
<tr>
<th>KD4 (Reference Malay dictionary)</th>
<th>IMED (Active Bilingual)</th>
<th>KPBMBI (Passive Bilingual)</th>
</tr>
</thead>
<tbody>
<tr>
<td>beri; memberi 1. menyampaikan (sesuatu) kpd seseorang (utk disimpan, digunakan atau utk sesuatu keperluan): kerani itu ~ saya dua helai borang kosong utk diisi; dia ~ adiknya sebuah buku cerita; 2. bp tidak menghalang seseorang melakukan sesuatu, membenarkan, mengizinkan, merelakan: saya tidak diberinya masuk; 3. mengakibatkan, menyebabkan, menjadikan: ~ malu; ~ bercat; 4. menunjukkan, membuka: ~ jalan; 5. = ~ bahagian bp memukul: awas kau, nanti aku beri; 6. kenyal, menganjal: tanah liat yg lembut, ~ dan menganjal; ~ bertenggang memberi kelonggaran (kesempatan) utk beriktiar; ~ hati a) memanjakan; b) memberikan hati; ~ hormat (salam, selamat); ~ ızin mengizinkan; ~ muka memanjakan; ~ telinga sujud mendengar (kan); ~ barang kpd tangan kera prb menyerahkan sesuatu kpd orang yg akan merosakannya;</td>
<td>beri, memberi 1. to give: Ibu ~ sepuluh sen kpd peminta sedekah itu. Mother gave ten sen to the beggar. 2. to result in; to cause: Wahab tak mahu ~ malu kawannya itu. Wahab does not want to embarrass his friend. 3. to allow, to permit: Dia ~ saya tempoh sebulan utk membayar hutang itu. He gave me a month to settle the debt. ~ hormat (salam) to extend regards; to greet. ~ hati to allow; to permit; ~ jalan to make way; to move aside (so that someone or something can pass through); ~ muka to spoil; to indulge.</td>
<td>beri = memberi kj 1. membagikan, menyerahkan (lwn ambil) to give, to hand over. Saya akan memberi hadiah itu kepada awak. 2. membenarkan, mengizinkan to allow; to permit. Kerajaan memberi peluang kepada kaum bumiputera menyertai kegiatan ekonomi negara. 3. mengakibatkan, menjadikan, menyebabkan to cause, to result in. Perbuatan Ah Kong itu telah memberi malu kepada keluarganya. ~ ıngat to remind, ~ ızin to allow, to permit. ~ kuasa to authorize. ~ malu insult ~ memberi give and take, compromise, ~ milik alienate. ~ sumpah to take oath. ~ tabek to salute.</td>
</tr>
</tbody>
</table>
Theme – *hadiah kain pelikat dan kain batik / gifts of pelikat clothes and batik sarongs*
Receiver – *penduduk kampung yang miskin / poor kampung folks*

The GIVING PHYSICAL THEME then was mapped to the English GIVING frame. The mapping of the Malay and English frames resulted in suitable English equivalents for ‘*memberi*’. Table 3 below shows the database of the semantic information on Malay GIVING PHYSICAL THEME, whereas Table 4 illustrates the valence pattern for the GIVING PHYSICAL THEME frame.

**TABLE 3**
The database of semantic information on Frame GIVING PHYSICAL THEME

<table>
<thead>
<tr>
<th>LU</th>
<th>Grammatical Info</th>
<th>The Donor presents a Theme (physical entity) to the Receiver.</th>
</tr>
</thead>
<tbody>
<tr>
<td>memberi</td>
<td>Transitif Verb</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 4**
The valence pattern for the Frame GIVING PHYSICAL THEME

<table>
<thead>
<tr>
<th>Donor NP Subject</th>
<th>Node</th>
<th>Theme NP Object</th>
<th>Receiver NP Complement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wakil Rakyat</td>
<td>memberi</td>
<td><em>hadiah kain pelikat dan kain batik</em></td>
<td>kepada penduduk kampung yang miskin</td>
</tr>
</tbody>
</table>

**DISCUSSION**

The frame-based analysis contributed to the semantic information of the Malay verb ‘*memberi*’. The frame-based analysis of ‘*memberi*’ identified five Malay frames: GIVING PHYSICAL THEME, GIVING ABSTRACT THEME a parent frame that includes PERMISSION and ATTENTION, COMMUNICATION, CAUSATION and PERMITTING. The comparison of ‘*memberi*’ from the three dictionaries and the frame-based analysis is displayed in Table 5 below.

The frame-based analysis revealed two additional frames and they are the frames ATTENTION from GIVING PHYSICAL ABSTRACT and the frame of COMMUNICATION. The additional frames contributed to finer distinctions of meanings for ‘*memberi*’ and the suitable equivalents. For example, Table 3 above shows the LUs or the English equivalents for ‘*memberi*’ from the frame GIVING PHYSICAL THEME.

The Proposed Malay Framenet

From the frame-based analysis shown above, this paper therefore proposes the possibility...
of building a Malay FrameNet based on the architecture of the English FrameNet, originally set up at the International Computer Science Institute in Berkeley. The semantic information shown earlier in Table 3 of the entry ‘memberi’ of the frame GIVING PHYSICAL THEME is an example of how the semantic information of Malay lexical is captured in the database. This proposed database must also provide the English translations of the example sentences as well as English lexical units or LU sourced from the English FrameNet.

There are in principles two different approaches in the effort to build a Malay FrameNet:

1. Reuse the existing corpus resources, and in this case, the 5 million words DBP-UKM Malay corpus. A full coverage of semantic annotation is needed to produce a reasonable database for Malay.

2. Build machine readable dictionaries (MRD) from scratch by involving experts and Malay linguists and also trained lexicographers. Fontenelle (2009) adds that to build a bilingual lexical-semantic database, the researcher needs to annotate frames manually. The researcher cum coder must be familiarised with the concepts of frames and the frame elements (FE). Therefore, training is needed for the coders involved.

### Table 5
The meanings of ‘memberi’ provided by the dictionaries as compared to the frames found in Malay after the FS analysis

<table>
<thead>
<tr>
<th>Meanings</th>
<th>KD4</th>
<th>IMED</th>
<th>KPBMPI</th>
<th>Frames in Malay (FS analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To present / provide something physical to someone</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>GIVING PHYSICAL THEME</td>
</tr>
<tr>
<td>2. To allow, give permission, not stopping anyone to do something</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>GIVING ABSTRACT THEME - PERMISSION</td>
</tr>
<tr>
<td>3. To give attention (abstract matter)</td>
<td></td>
<td></td>
<td></td>
<td>ATTENTION</td>
</tr>
<tr>
<td>3. To cause something</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>CAUSATION</td>
</tr>
<tr>
<td>4. To show the way, to give way</td>
<td></td>
<td></td>
<td></td>
<td>PERMITTING</td>
</tr>
<tr>
<td>5. For communication purposes</td>
<td></td>
<td></td>
<td></td>
<td>COMMUNICATION</td>
</tr>
</tbody>
</table>
Malay Framenet: An Application for Bilingual Dictionaries

The setting up of the Malay FrameNet can be seen as an effort towards processing the Malay language and developing the Malay computational lexicography. It is important to encourage future linguistic and computational lexicography research to organise joint efforts in researching the Malay language, digitising the results as well as promoting natural language processing of the Malay language.

CONCLUSION
This paper discussed the transformation and innovation in bilingual dictionary research by building up a database. This paper also presented a short discussion on matters pertaining to the Berkeley or the English FrameNet, FrameNets of other languages, and the theory of Frame Semantics. This paper then further discussed how the frame-based analysis was carried out on the 5 million word DBP-UKM Malay corpus on the lexical ‘memberi’ and how the analysis contributed to the semantic information of the proposed Malay FrameNet.

ACKNOWLEDGEMENT
This article is based on a research project (Code: GGPM-2011-048) funded by Universiti Kebangsaan Malaysia.

REFERENCES


