

5th World Conference on Educational Sciences - WCES 2013

The use of corpus and Frame Semantics in a lexicography class: Evaluating dictionary entries

Intan Safinaz Zainudin *, Nor Hashimah Jalaluddin, Khairul Taufiq Abu Bakar

*School of Language Studies and Linguistics,
Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia*

Abstract

In a lexicography class, linguistics students are taught the principles of arranging dictionary entries with the application of theories. There is also a need for the students to evaluate existing dictionary entries as well. Due to the importance of corpus use in a lexicography work (Atkins & Rundell, 2008), students were introduced to a Malay corpus and the Frame Semantics (Fillmore et al., 2003). Students were taught how to analyse the meaning of a word based on the frames and frame elements. The use of DBP-UKM 5 million Malay corpus and the framework of Frame Semantics were found to be beneficial in assisting students to evaluate the verb entries of *memerlukan* and *mencari*, two highly frequent verbs in Malay. Students were able to evaluate the two verb entries from the Kamus Dewan 4th Edition (KD4), a monolingual dictionary, and the bilingual Intelek Malay-English Dictionary (IMED). Students' analysis findings will be presented and a few recommendations on teaching lexicography will be discussed in this paper.

© 2013 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Selection and/or peer-review under responsibility of Academic World Education and Research Center.

Keywords: Corpus, Frame Semantics, lexicography class, evaluating, dictionary entries.

1. Introduction

Lexicography is a professional activity and academic field concerned with dictionaries and other reference works. Hartmann (2001) further discusses the difference between the practice of lexicography that involves dictionary making and the theory that involves dictionary research. The subject of lexicography that is usually taught at the graduate level in a university mainly focuses on academic aspects related to dictionaries. Students are taught the principles of writing and compiling of dictionary entries along with different linguistic theories that can assist in compiling and editing entries. This is in line with the most essential feature of lifelong learning that leads toward the improvement of an individual and the betterment of society at large (Anuwar, 2005).

In the Malaysian context, research on bilingual dictionaries primarily focuses on the problems and the lack of equivalents for entries (Noresah & Fadilah, 2005). However, not much focus is given to the linguistic information of bilingual dictionary entries based on corpus analysis. Given this situation, the focus of this paper is to look at how the theory of Frame Semantics is taught to students and how they are trained to edit and upgrade bilingual dictionary entries supported with corpus data. A dictionary is considered 'reliable' if it provides evidence on its use and usage of each entry. Atkins & Rundell (1998) argue for the importance of lexicographic evidence in the pre-lexicography

* Corresponding Author name. Intan Safinaz Zainudin Tel.: +6-019-385-0011

E-mail address: intansz@ukm.my

compilation process. Lexicographic evidence is defined as citations that are needed to support the use and meaning of each entry. Corpus data can provide lexicographic evidence in dictionary making, due to advances in computer software and the Internet. Corpus is a systematic collection of texts in digital form that documents the usage features of a language and language variety. According to Hartmann & James (1998), the practical uses of computers for processing data and the theoretical advances of Corpus Linguistics have provided lexicographers powerful tools for storage and retrieval of data to describe all aspects of language. This is evident with the vast development and production of dictionaries, for example, the production of the Macmillan English dictionary that is supported with linguistic information from the British National Corpus.

2. Theoretical framework

This paper shows the importance of teaching Frame Semantics and FrameNet in finding meaning discrimination of two verb entries *mencari* (to look for) and *memerlukan* (need) from a bilingual dictionary *Intelek Malay-English Dictionary* (IMED). The two verb entries are chosen as they are among two of the highest frequently used Malay verbs in the corpus (Intan Safinaz Zainudin, 2011).

2.1 Frame Semantics

Frame Semantics (henceforth FS) is a theory ‘that links the meanings of words explicitly to the syntactic contexts in which the words occur’ (Atkins, Fillmore & Johnson, 2003: 254). Frames are seen as concepts that are systematic and inter-connected. To understand a concept we must understand the systematic concepts that are rendered automatically in a text, or a conversation (Fillmore, 1982). Fillmore also introduced the ‘valency theory,’ which is a description of a lexical and how the frame elements may be expressed in a pattern of grammatical constructions in its context (Atkins, Fillmore & Johnson, 2003).

2.2 FrameNet

FrameNet is an online English language lexical database. FrameNet is based on the principles of FS and can be accessed at <https://framenet.icsi.berkeley.edu>. The FrameNet project was set up at the International Computer Science Institute in Berkeley, California with a database that describes several thousand words of the 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 utilized to support the meaning of each frame and its elements.

3. Methodology

In class, students are taught the theory of FS and how the theory helps to provide a complete description of a lexical item. FrameNet is an English language database that provides information only on English frames and the semantic and syntactic information. Therefore, students must be taught to apply FS to the Malay corpus. Each student is given two bilingual dictionary entries to work on. In this paper two verbs are chosen which are *mencari* (to look for) and *memerlukan* (need) from the *Intelek Malay-English Dictionary* (IMED). For the purposes of this article, the DBP-UKM 5 million Malay corpus is utilised. A team of researchers from DBP and the Universiti Kebangsaan Malaysia developed the DBP-UKM 5 million corpus (Zaharani, 2007). The 5 million corpus was extracted from the then 100 million DBP corpus. The extraction of the corpus was based on the principle that the smaller corpus must be a balanced and representative corpus from the bigger 100 million corpus. The 5 million corpus consists of sub corpora of magazines, books and newspapers.

The following steps are taught to students on how the FS approach can be adopted for the Malay corpus. The FS analysis supported by corpus data is discussed below.

1. Compare the semantic description of a Malay entry from the chosen dictionaries.
2. Refer to the corpus data. Note the most frequent constructions.

3. Identify the frame and frame elements.
4. Identify the valence group.
5. Identify the description of the frame, the frame elements, and example Malay sentences suitable to be suggested for the dictionary entry.
6. Identify the matching English frame from FrameNet that can be mapped on to the Malay.
7. Source the English equivalents from the mapped English frames.
8. Translate the Malay corpus sentences into English as example sentences.

4. Discussion

From the students' work it is evident that they are able to apply FS analysis to the 5 million UKM-DBP Malay corpus. Table 1 and Table 2 present a complete meaning description for frame MENCARI FIZIKAL and frame MENJADIKAN PERLU respectively. Frame MENCARI FIZIKAL is the core meaning for the entry *mencari* as it is recorded in KD4 and IMED. Table 1 provides the complete semantic and syntactic descriptions that include the frame meaning, frame elements, a Malay sentence example from the corpus, English translation and suggested lexical units that can be recommended as English equivalents.

Table 1: The meaning description of the frame MENCARI FIZIKAL, frame elements, valence patterns and other linguistic information

LEMMA	Cari
Lexical Unit (LU)	Mencari
Word class	Verb
GIVING PHYSICAL frame	An Agent is looking for an Entity of a physical entity.
Frame Elements	Agent, Entity
Example Sentence	<i>Ketiga-tiga ahli pasukan pandu puteri itu pun pergi mencari kayu api.</i>
The valence patterns	Transitive (Agent/NP/Subject) + <i>mencari</i> + (Theme//NP/Object)
English translation	The three members of the girl scout team went looking for firewood.
English Lus	Forage, look, search, seek

Similarly, frame MENJADIKAN PERLU is the core meaning for the entry *memerlukan* as is recorded in KD4 and IMED. Table 2 provides the complete semantic and syntactic descriptions that include the frame meaning, frame elements, a Malay sentence example from the corpus, English translation and suggested lexical units that can be recommended as English equivalents.

Table 2: The meaning description of the frame MENJADIKAN PERLU, frame elements, valence patterns and other linguistic information

LEMMA	Perlu
Lexical Unit (LU)	Memerlukan
Word class	Verb
GIVING PHYSICAL frame	Kebergantungan pada sesuatu perkara adalah berlandaskan kepada ketetapan tertentu bagi mendapatkannya ataupun merealisasikannya.
Frame Elements	Dependence, Regulation
Example Sentence	<i>Bagaimanapun, usaha sama itu memerlukan kelulusan kementerian pertanian dan kesihatan Malaysia serta Jabatan Agama Islam sebelum pelaksanaan sebarang projek, katanya.</i>
The valence patterns	Transitive (Dependence/NP/Subject) + <i>memerlukan</i> + (Regulation//NP/Object)
English translation	However, the joint venture requires the approval of the Ministry of Agriculture and Health Malaysia and the Islamic Religious Department before implementation of any project, he said.
English Lus	needed, required

Further comparison between the entries of KD4, IMED and the FS analysis also highlight important issues. As shown in Table 3 and 4, not all meanings provided in the reference dictionary KD4 can be analysed with FS due to the unavailability of attested corpus data. For example, meaning 2 for *mencari* shown in Table 3 is not found in the

FS analysis, as there is no corpus sentence to support meaning 2. Similarly for *memerlukan*, meaning 3 is not found in the FS analysis. The FS analysis has successfully provided comprehensive meaning descriptions for meanings that are only available in the 5 million corpus. The corpus in use for dictionary research should be big and comprehensive enough, more than 5 million words in order to provide an adequate reference for meaning as in the discussion of how big a corpus should be (Atkins & Rundell, 2008) along with linguistic information for bilingual dictionary entries (Royle, Richardson & Bourguignon, 2009).

Table 3: The meanings provided in the two dictionaries and the FS analysis for *mencari*

Meaning	KD4	IMED	FS Analysis
1. to get something	√	√	√
2. to create something	√		

Table 4: The meanings provided in the two dictionaries and the FS analysis for *memerlukan*

Meanings	KD4	IMED	FS Analysis
1. menjadikan sesuatu itu wajib dilakukan	√	√	√
2. menghendaki sangat-sangat	√	√	√
3. membuat sesuatu kerana memikir lebih perlu daripada perkara lain	√		

5. Conclusion

The students' analyses has provided some insightful findings suggesting Frame Semantics as a suitable theory to assist lexicographers in providing comprehensive analysis on bilingual dictionary entries. The practical and hands-on FS analysis, using corpus as attested data, provides good training for graduate linguistic students. However, the UKM-DBP 5 million corpus data of 5 million is not big enough to provide evidence for certain meanings for *mencari* and *memerlukan*. A bigger Malay corpus is needed to provide more usage of the meanings recorded in the dictionaries.

Acknowledgements

This article is based on research projects GGPM-2011-048 and PTS 2012-032 funded by the Universiti Kebangsaan Malaysia.

References

- Anuar Ali. (2005). *Lifelong learning: Policy of Malaysia and the role of OUM*. Seminar organised by Korea National Open University.
- Atkins, B.T.S. Fillmore, C.J. & Johnson, C.R. (2003). Lexicographic relevance: Selecting information from corpus evidence. *International Journal of Lexicography*, 16 (3): 251-280.
- Atkins, B.T.S. Rundell, M. & Sato, H. (2003). The contribution of FrameNet to Practical Lexicography. *International Journal of Lexicography*, 16 (3): 333-357.
- Atkins, B.T.S. & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.
- Fadilah Jasmani, (2005). Kamus dwibahasa: Terjemahan lengkap dan separa lengkap. *Prosiding CD-Rom Persidangan Antarabangsa Leksikologi dan Leksikografi Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Fillmore, C. J. (1982). Frame Semantics. In The Linguistics Society of Korea (Ed.) *Linguistics in the morning calm*. Seoul: Hanshin Publishing Co.
- Fillmore, C.J. Johnson, C.R. & Petruck, M.R.L. (2003). Background to FrameNet. *International Journal of Lexicography*, 16 (3): 235-250.
- Hartmann, R.R.K. (2001). *Teaching and researching lexicography*. Pearson Education limited: London.
- Hartmann, R.R.K. & James, G. (1998). *Dictionary of lexicography*. Routledge: London.
- Ibrahim Ahmad. (2005). *Perkamus Melayu teori dan amali*. Kuala Lumpur: Penerbit Universiti Malaya.
- Intan Safinaz Zainudin. (2011). Analisis makna dan padanan dalam kamus dwibahasa Melayu-Inggeris. Unpublished PhD thesis. Universiti Kebangsaan Malaysia.
- Kamus Dewan. Edisi Keempat. (2005). Kuala Lumpur: Dewan Bahasa dan Pustaka.

- Kamus Intelek Malay-English Dictionary*. (1996). Kuala Lumpur: Arus Intelek Sdn Bhd.
- Noresah Baharom& Fadilah Jasmani. (2003). Penterjemahan dalam penyusunan kamus dwibahasa pengalaman DBP. *Persidangan Penterjemahan Antarabangsa ke-9. 2-4 September 2003*. Dewan Bahasa dan Pustaka.
- Royle, P., Richardson, I., Bourguignon, S. B. N. (2009). Is bigger better? Corpus and dictionary use in the search for compounds, collocations, derived forms and fixed expressions. *Meta: Translators' Journal: 54 (3): 520-532*.
- Zaharani Ahmad. (2007). *Aspek nahu praktis Bahasa Melayu*. Bangi: Penerbit UKM.