

ICLICE 2015 42 Panornuang Sudasna Na Ayudhya

## WordNet Development of English-Thai-Lao-Vietnamese 1<sup>st</sup> Order Entity Words

Panornuang Sudasna Na Ayudhya  
Research and Development Institute  
Bansomdejchaopraya Rajabhat University, Bangkok, Thailand  
E-mail address: panor\_sudas@bangkokmail.com

### ABSTRACT

WordNet is a kind of lexical database which is well-known and has the influence on many computational linguistic related applications. The purposes of this research were to examine the equivalent translation of English-Thai-Lao-Vietnamese words in the 1st Order Entity using bi-directional translation method and to develop WordNet from the selected equivalent translation of English-Thai-Lao-Vietnamese words in the 1st Order Entity. The equivalent translation of English-Thai-Lao-Vietnamese words in the 1st Order Entity was examined by bi-directional translation method using the native speakers of English, Thai, Lao, and Vietnamese as the key informants. They were asked to translate from source language to target languages and then from target to source language; for example, from English to Thai and from Thai to English. The bi-directional correctness was calculated using F-Measure in order to select the equivalent translation of English-Thai-Lao-Vietnamese words in the 1st Order Entity. The translation equivalent pairs of English-Thai-Lao-Vietnamese in the 1st Order Entity were selected. The English-Thai-Lao-Vietnamese WordNet was developed from the selected equivalent translation 1st Order Entity words.

*Keywords:* WordNet, English-Thai-Lao-Vietnamese words, the 1st Order Entity, Equivalent Translation

### Introduction

WordNet is a widely used lexical database that groups words into sets of synonyms and categorizes them in four categories of noun, verb, adjective, and adverb. WordNet has been developed since 1985 by a group of psychologists and linguists at Princeton University. In the present, the importance of a lexicon including the phonological, syntactic, and lexical components for linguistic production and comprehension has been increased. The integration of these components in the lexicon has been incorporated into the psycholinguistic and lexicography aspects or psycholexicology (Miller 1985).

Thus, WordNet is the result obtained from the development of lexicon based on psycholexicology assumption. WordNet is different from a standard dictionary because words in WordNet are linked together by their semantic relationships (Garrett 1982).

The present research was interested to develop WordNet using translation equivalence of English-Thai-Lao-Vietnamese words in the 1st Order Entity. In this research report, the examination of equivalent translation of English-Thai-Lao-Vietnamese words in the 1st Order Entity using bi-directional translation method and the development of WordNet from the selected equivalent translation of English-Thai-Lao-Vietnamese words in the 1st Order Entity were presented.

In the following sections, the notion of WordNet, the four orders of entity, and feature of translation equivalence will be briefly mentioned as the study background. Then, the aim of the study, research objectives, and research methodology will be presented. Finally, the results of the study will be summarized and discussed in the end of paper.

## Theoretical Background Statement of the Problem

### WordNet

In Natural Language Processing, WordNet is a well-known lexical database that groups words into sets of synonyms and categorizes them in four categories: noun, verb, adjective, and adverb (Miller 1985, 1986). This categorization of WordNet was supported from psycholexicology that syntactic categories differ in subjective organization emerged first from studies of word associations. The most important feature of WordNet is organizing lexical information in terms of word meanings, rather than word forms.

### Four Orders of Entity

Lyons (1977) presented three orders of entities, which refines the traditional distinction between concrete and abstract nouns. The three orders of entities are defined as following.

**First-order entity** is defined as physical objects such as persons, animals, and things. First-order entity is evaluated in terms of their existence.

**Second-order entity** is mentioned as events, processes, states-of-affairs, etc., which are located in time. Second-order entity is evaluated in terms of their reality.

**Third-order entity** is mentioned as abstract entities, which are not located in space and time. Third-order entity is evaluated in terms of their truth.

Later, Hengeveld (1992) proposed a fourth order entity, which is located in space and time, and is evaluated in terms of their felicity.

For the present research, only the 1st Order Entity of words was included in the investigation.

### Translation Equivalence

Translation equivalence is a linguistic principal concept in translation theory. It is a constitutive feature and the guiding principle of translation. The notion of translation equivalence has been elaborated in many translation theories since the twentieth century. However, the notion of translation equivalence is still one of the most controversial areas in the field of translation theory. The different kinds of equivalence was described by many theorists as Vinay and Darbelnet (1958), Jakobson (1959), Nida and Taber (1969), Catford (1965), House (1997), Koller (1979), Newmark (1981), Baker (1992), and finally, Pym (2010). In the following, the concept of equivalence in translation obtained from each notion will be summarized briefly.

**Vinay and Darbelnet.** Vinay and Darbelnet (1958) distinguish between *direct* and *oblique* translation. According to Vinay and Darbelnet (1958), *direct* translation refers to literal translation and *oblique* translation refers to free translation. They also proposed seven translational procedures: borrowing, calque, literal translation, transposition, modulation, equivalence and adaptation, the first three covered by direct translation and the remaining four by oblique translation.

**Jakobson.** The structuralist Roman Jakobson (1959) proposed three kinds of translation: *intralingual* (rewording or paraphrasing within one language), *interlingual* (rewording or paraphrasing between two languages), and *intersemiotic* (rewording or paraphrasing between sign systems). According to the translation equivalence, he presented that there is no full equivalence between two words of two languages.

**Nida and Taber.** Nida and Taber (1969) presented that there are two basic types of equivalence: (1) *formal equivalence* and (2) *dynamic equivalence*. In formal equivalence, the target language resembles the source language in both form and content; whereas, in dynamic equivalence an effort is made to convey the source language in the target language as naturally as possible.

**Catford.** Catford (1965) distinguished between *types* and *shifts* of translation. *Shifts* were defined as the changes that take place during the translation process. For the types of translation, there are three types of translation. Firstly, *full translation* is contrasted with *partial translation* according to the extent of translation. Secondly, *total translation* differs from *restricted translation* according to the levels of language involved in translation, and thirdly, Catford distinguished between *rank-bound translation* and *unbounded translation* depending on the grammatical or phonological rank at which equivalence is established.

**House.** House (1997) distinguished between two basic types of translation, *overt translation* and *covert translation*. An overt translation referred to a target language that consists of elements that it is a translation. On the other hand, a covert translation is a target language that has the same function with the source language.

**Koller.** According to Koller (1979), equivalence deals with equivalent items in source language - target language pairs and contexts. Koller (1979) proposed five types of equivalence: (a) denotative equivalence involving the extralinguistic content of a text, (b) connotative equivalence relating to lexical choices, (c) text-normative equivalence relating to text-types, (d) pragmatic equivalence involving the receiver of the text, and, (e) formal equivalence relating to the form and aesthetics of the text.

**Newmark.** Newmark (1981) proposed *semantic* and *communicative translation*. The semantic translation focuses on meaning whereas communicative translation concentrates on effect. It should be pointed out that during the translation process, the two methods of translation may be used in parallel.

**Baker.** Baker (1992) proposed that equivalence is a relative notion because it is affected by a various linguistic and cultural factors. Baker distinguished 3 types of translation equivalence: Grammatical equivalence, textual equivalence, and pragmatic equivalence. *Grammatical equivalence* refers to the diversity of grammatical categories across languages and the difficulty of finding an equivalent term in the target language because of different grammatical rules across languages. *Textual equivalence* refers to equivalence between a source language and target language in terms of cohesion and information. Finally, *pragmatic equivalence* focuses on implicature.

**Pym.** Pym (2010) distinguished between *natural* and *directional equivalence*. Natural equivalence exists between languages prior to the act of translating, and, secondly, it is not affected by directionality. The most important assumption of directional equivalence is that it involves asymmetry in translation.

The concept of translation equivalence applied for the present study was the notion proposed by Newmark (1981). Newmark (1981) proposed *semantic* and *communicative translation*. The present study will focus on the type of *semantic translation equivalence*.

### Research Purposes

1. To examine and select the translation equivalence of English-Thai-Lao-Vietnamese words in the 1st Order Entity using bi-directional translation method.
2. To develop English-Thai-Lao-Vietnamese WordNet.

### Research Methodology

The research methodology was presented as two subsections for the methodology used to examine and select the translation equivalence of English-Thai-Lao-Vietnamese words in the 1st Order Entity using bi-directional translation method and for to develop English-Thai-Lao-Vietnamese WordNet.

### **Examining and Selecting the Translation Equivalence of English-Thai-Lao-Vietnamese Words in the 1st Order Entity using Bi-Directional Translation Method**

The 1st Order Entity words were selected from Brown Corpus (Word frequency corpus)

The equivalent translation of English-Thai-Lao-Vietnamese words in the 1st Order Entity was developed by following translational procedure:

a. The two English (Native language, henceforth, NL) -Thai (Foreign language, henceforth, FL) henceforth), two Thai (NL) -English (FL), two English (NL) -Lao (FL), two Lao (NL) - English (FL), two English (NL) - Vietnamese (FL), and two Vietnamese (NL) - English (FL) bilinguals were assigned as the translators.

b. The translators independently translated the items bi-directionally and then the results were compared to obtain the most equivalent item. The bi-directional translation procedure included the process of the native speakers of each language were asked to translate from source language to target language and then back-translate from target language to source language. For instance, Thai native speakers who know English language were asked to translate from English to Thai and from Thai to English and also Lao native speakers who know English language were asked to translate from English to Lao and from Lao to English.

The translation results were tested using F-Measure ( $\geq 70\%$ ) (Shamsfard 2008).

### **Developing English-Thai-Lao-Vietnamese WordNet**

The words and meanings were selected and examined as the lexical substance of WordNet.

The source files that contain the lexical data and synsets were prepared by mapping English WordNet synsets to the translation equivalence words in Thai-Lao-Vietnamese. For instance, the synset of English word “adult” with the meaning of “a fully developed person from maturity onward” was mapped with Thai word “ผู้ใหญ่”, Lao word “ຜູ້ໃຫຍ່” and “ຄົນໃຫຍ່”, and Vietnamese word “người lớn.”

After the English synsets were mapped with Thai, Lao, and Vietnamese words, the obtained synsets of Thai, Lao, and Vietnamese words were interlinked by identical semantic relations as in WordNet.

The computer programs were developed to run the source files and synsets.

The developed WordNet was utilized offline. The English, Thai, Lao, and Vietnamese

### **Results**

290 equivalent translations of English-Thai-Lao-Vietnamese words in the 1st Order Entity have been selected and have been evaluated manually using F-Measure ( $\geq 70\%$ ). The examples of equivalent translation and the developed WordNet of English-Thai-Lao-Vietnamese words in the 1st Order Entity were shown in Table 1 and Figure 1.



This paper presented the inventory of selected 290 equivalent translation of English-Thai-Lao-Vietnamese words in the 1st Order Entity. WordNet of English-Thai-Lao-Vietnamese languages was developed using the corpus linguistics methodology. Eventually, there is a few numbers of equivalent translation of English-Thai-Lao-Vietnamese words in the 1st Order Entity which could get  $\geq 70\%$  of F-Measure. The results implied that evaluating the equivalent translation between words in more than two languages which are absolutely identical in meaning and usage using F-Measure is quite difficult to get a large number of equivalent translation of words in more than two languages. Hence, for the study of equivalent translation of many languages, it is proposed that equivalence between the source language and the target language should be studied based on the research purposes and must be identified clearly in term of which translation equivalence must be concerned.

### Acknowledgement

This research was supported by National Research Council of Thailand, Budget Year 2014. We also thank to all of native speakers of Lao and Vietnamese who were Master Degree students at Bansomdejchaopraya Rajabhat University, Thailand.

### References

- Baker, M. (1992). *In Other Words. A Coursebook on Translation*. London: Routledge.
- Catford, J.C. (1965). *A Linguistic Theory of Translation*. London: Oxford University Press.
- Hengeveld, K. (1992). 'Parts of speech' in Fortescue, M., Harder, P. & Kristoffersen, L. (eds.) *Layered structure and reference in a functional perspective*. Amsterdam: Benjamins. 29-55.
- House, J. (1997). *Translation Quality Assessment: A Model Revisited*. Tübingen: Narr.
- Jakobson, R. (1959). *On Linguistics Aspects of Translation*. In Venuti, L. (ed.) 2000, *The Translation Studies Reader*. London and New York: Routledge, 113-118.
- Garrett, M. F. (1982). 'Production of Speech: Observations from Normal and Pathological Language Use' in A. Ellis (ed.) *Normality and Pathology in Cognitive Functions*. London: Academic Press.
- Koller, W. (1979). *Einführung in die Übersetzungswissenschaft*. Heidelberg: Quelle and Meyer.
- Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.
- Miller, G. A. (1985). 'Wordnet: A Dictionary Browser' in *Information in Data, Proceedings of the First Conference of the UW Centre for the New Oxford Dictionary*. Waterloo, Canada: University of Waterloo.
- Miller, G. A. (1986). 'Dictionaries in the Mind.' *Language and Cognitive Processes*, 1, 171-185.
- Newmark, P. (1981). *Approaches to Translation*. Oxford and New York: Pergamon Press.
- Nida, E. and Taber, C.R. (1969). *The Theory and Practice of Translation*. Leiden: E.J. Brill.
- Pym, A. (2010). *Exploring Translation Theories*. London and New York: Routledge.
- Vinay, J. P. and Darbelnet, J. (1958). *Stylistique Comparée du Français et de l'Anglais: Méthode de Traduction*. Paris: Didier.