

---

# Semantic Role Labelling in Bilingual English-Vietnamese Corpus

Huynh Quang Duc<sup>1, \*</sup>, Tran Le Tam Linh<sup>2</sup>

<sup>1</sup>Computer Science Center of Soc Trang Vocational College, Soc Trang Province, Vietnam

<sup>2</sup>Center for Mathematical Science, University of Science, National University, Ho Chi Minh City, Vietnam

## Abstract

Issue about Semantic Role Labelling (SRL) for bilingual has been studying on many popular languages as English, French, etc. However, Semantic Role Labelling tasks for unpopular languages as Vietnamese are currently limited, especially for making the most of semantic similarities on bilingual English-Vietnamese. In this paper, we propose a solution for Semantic Role Labelling tasks automatically on bilingual English-Vietnamese Corpus in order to take full advantages of the translations of cross-language lexicalization, but it also maintains the core elements of its semantic role. This system has used corpus from the Web to build sets associated with the ability to combine many different meaning words found in the corpus, and it has also used an unsupervised algorithm to label the semantic role in English, which based on semantic similarities through English-Vietnamese corpus. Then, this system will automatically project labels from English to Vietnamese via available links.

## Keywords

Unsupervised, Bilingual, Parallel Corpus, Semantic Role Labelling, Machine Translation

Received: March 31, 2015 / Accepted: April 17, 2015 / Published online: June 26, 2015

@ 2015 The Authors. Published by American Institute of Science. This Open Access article is under the CC BY-NC license.

<http://creativecommons.org/licenses/by-nc/4.0/>

---

## 1. Introduction

Semantic Role Labelling (SRL) system has played an important role as a tool processing natural language, especially, in the period of extremely rapid development of data on the Internet. Nowadays, the biggest problem which many scientists as well as linguists are focusing to resolve is how to reduce ambiguity in natural language to help computers that can be understand meaning of words in human speech in different fields such as information retrieval, question answering, summarization, machine translation etc.

In fact, The sentence-level semantic analysis of text is concerned with the characterization of events, such as determining “who” did “what” to “whom,” “where,” “when,” and “how.” The primary task of semantic role labelling (SRL) is to indicate exactly what semantic relations hold

among a predicate and its associated participants and properties, with these relations from a pre-specified list of possible semantic roles for that predicate (or class of predicates) (Marquez, Carreras, Litkowski, Stevenson, 2008).

Besides, there are some vital factors including learning machine technology, widespread of sense label system in Word Net and availability of large corpus have been interested in word sense disambiguous. Mainly, supervised systems which learn from correctly semantic role labelled corpus that is manually made by linguistic experts. However, learning to evaluate on training corpus needs a large labelled data (Mona Diab and Philip Resnik, 2002). This affair is very expensive in cost and consumes a lot of time, which requires a professional team about labelling semantic language. Unsupervised methods have the advantage of making fewer assumptions about availability of data, but ability to lower general in practice (Philip Resnik, 1997; Lin, 2000).

---

\* Corresponding author

E-mail address: [dhuyhquang@gmail.com](mailto:dhuyhquang@gmail.com) (H. Q. Duc)

Moreover, according to Gildea and Jurafsky (2002), the ability to semantic role label automatically on most data of parallel corpus can be performed by an unsupervised algorithm thanks to its reasonable cost and less time.

In this paper, upon using parallel corpus, we take the advantages of two languages to exploit reasonably. Also, we simultaneously use the role semantic label available on bilingual English-Vietnamese corpus. The method aims at achieving two main goals: Firstly, producing some data that is labelled semantic role on English with semantic inventory which is unnecessary to be manually made by experts. Secondly, achieving semantic role labelling that is the same semantic inventory for Vietnamese.

Important problem of this study is the observation of the translation which can be met reciprocity as a basis feature in semantic role (Nancy Ide, 2000). One word that has multiple senses in English is often translated as distinct words in Vietnamese, with the particular choice depending on the translator and the contextualized meaning. So, an appropriate translation is seen as a semantic indication for an example in its context. On the other hand, that handful of words is rarely a singleton set even for a single word/sense, because the preferences of different translators and the demands of context produce semantically similar words that differ in their nuances.

For example, in an English-Vietnamese parallel corpus, the Vietnamese *đường* could be found in correspondence to English *road* in one instance, and to *sugar* in another. But we can take advantages in practice that two words in English to appearance correspondence with word “*đường*” in Vietnamese to predict two word English have some specific factors about meaning in particular paragraphs. We can use those predictions to determine the meaning of English words that is mentioned, which is concordant with initial target so that we can project a semantic choice of word “*đường*” in Vietnamese to “*road*” or “*sugar*” in English. Thus, semantic role labelling in parallel languages with single semantic inventory is entirely consistent and ability performed.

The purpose of this study is that we have built the English-Vietnamese parallel corpus with the majority of the data taken from the internet. Then we will conduct to determine the semantic role of the noun in the bilingual sentence pairs of English-Vietnamese languages via their semantic similarities and translation of cross-language lexicalizations to identify semantic role labels for English nouns. And in the last step, we will project the semantic role labels on English to the corresponding words in Vietnamese via the word level alignment. In fact, the set of label which is used for labelling in this article belongs to the bilingual concept dictionary LLOCE (Longman Lexicon of Contemporary English) -

LLOCV (Longman Lexical of Contemporary Vietnamese).

The remains of this paper are as follows:

- Proposed approach method: Describe contents of performance to semantic role labelling in parallel English-Vietnamese corpus.

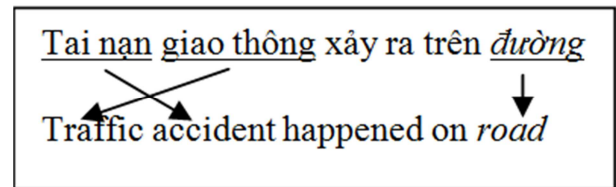


Figure 1. An example for a noun-aligned.

- Evaluated approach method: Present necessary requirements in evaluating experiment results and resources that we use for semantic role label.
- Discussion about issues we take advantages in parallel corpus.
- Conclusion and future development.

## 2. Proposed Approach Method

For convenience in approach of research method, in parallel English - Vietnamese corpus, we can count the semantic statistic of English. Although there is no necessary assumption of directionality in translation, we will refer to the English language corpus as the target language to semantic role label and the Vietnamese language corpus as the source corpus, which corresponds to the characterization. In the previous section, our example is word “*đường*” translated into two different words in English such as “*road*” and “*sugar*” in two different contexts. The process can be described more details for an approach method as follow:

1. Identify words in the target (English) corpus and their correspondence in the source (Vietnamese).

For example in this case, we have an ability set in English corpus {*road*, *sugar*} and a word in Vietnamese corpus {*đường*}.

We suppose a sentence or a paragraph in corpus that is translated parallel. Parallel data are available for bilingual English-Vietnamese corpus via the Web on Internet. After identifying and tokenizing sentences with words that can be associated, we will obtain word-level alignments for the parallel corpus using the GIZA++ model. For each word in Vietnamese instance  $w$ , we collect a word instance  $v$  that it's aligned. Then, positions of that word in the example are recorded so that in the following section we can project eventual semantic role labels from  $v$  to  $w$ . For another example, we have aligned a couple of bilingual English-

Vietnamese sentences as figure 1: alignments can occur between the word “đường” and “road” in couple of bilingual sentences of figure 1, meaning the system will translate “đường” to “road”, “tai nạn” to “accident” and “giao thông” to “traffic”.

2. Group the words of the target language – forming ability sets – that were translated into the same orthographic form in the source language (Vietnamese). For instance, we use corpus to build all ability sets of words that can be aligned with many words (two or more words) which are detected in parallel corpus. We collect for each type of word  $v_i$  in Vietnamese that includes all the type of words in English

which are aligned anywhere in the corpus that we call the ability set of  $v_i$ . For another example in this case, we have word “đường” in Vietnamese can be included the type of words in English such as *road, sugar, line*. We have the word *line* added in the ability set because in some other cases in parallel corpus that “đường này hơi tối” is translated into “this line is a bit dark”. Moreover, in the ability set can be included more other words if the system detects in English-Vietnamese corpus that those sentences have alignments which can be translated word “đường” into the other word in English (see figure 2).

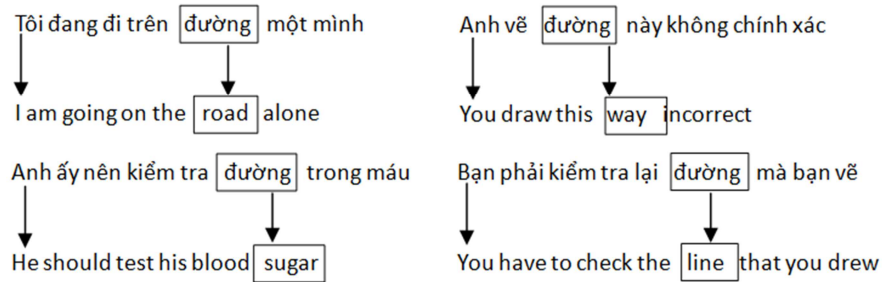


Figure 2. An example about building ability sets.

With the pair of parallel sentences in figure 2, the result of the ability set in English can be created as  $\{road, way, sugar, line\}$  from the source set in Vietnamese  $\{đường\}$ .

The contents in the step 1, 2 can be described as some basis steps by algorithm as figure 3 as follow:

**Algorithm 1:**  
 Step 1: Find nouns  $v_i$  (Vietnamese) in parallel corpus that was aligned and build Vietnamese noun set; we call the candidate set  $V$ .  
 Step 2: Create empty set  $S$  to record the ability sets for each word  $v_i$ ;  
 Step 3:  $k = 1$ .  
 Step 4: for each word  $v_i$  in  $V$   
     Create empty  $S_k$ ;  
     For each sentence pair in parallel corpus  
         Find word pair  $(w_v, w_e)$  that having  $w_v = v_i$ , adding word English into  $S_k$ ;  
         Record its link;  
 Step 5: Add  $(v_i, S_k)$  into  $S$ .  
 Step 6:  $k = k + 1$ .  
 Step 7: Repeat from step 4.

Figure 3. The basic Steps of Algorithm for building Ability Sets.

3. Within each of the ability sets, consider all the possible semantic labels for each word and select semantic labels informed by semantic similarity with the other words in the group. For example, as in within the ability set  $\{road, sugar, line\}$  and the source set  $\{đường\}$ , we will consider the pairs  $(road, đường)$ ,  $(sugar, đường)$ ,  $(line, đường)$ , whose pairs

will be assigned a confidence of its sense. In this step, the ability set will be considered as an issue of semantic role label on monolingual toward semantic inventory on the target language. We consider the ability set  $\{road, sugar, line\}$ , for human, choosing these semantic words in the ability set is very simple, but for computer, determining the meaning of words performed is through statistics by calculating a probability algorithm. We use the idea that is exploited by Resnik’ algorithm for disambiguating groups of related nouns (Philip Resnik, 1999b). Besides, we also refer to the approach of Resnik (Philip Resnik, 1997) about *selectional* reference and sense disambiguation. His model defines the *selectional* preference strength of a predicate as:

$$S_R(p) = D(\Pr(c | p) || \Pr(c)) = \sum_c \Pr(c | p) * \log \frac{\Pr(c | p)}{\Pr(c)}$$

Intuitively,  $S_R(p)$  measures how much information, in bits, predicate  $p$  provides about the conceptual class of its argument. The better  $\Pr(c)$  approximates  $\Pr(c | p)$ , the less influence  $p$  is having on its argument, and therefore the less strong its *selectional* preference. Given this definition, a natural way to characterize the "semantic fit" of a particular class as the argument to a predicate is by its relative contribution to the overall *selectional* preference strength. In particular, classes that fit very well can be expected to have higher posterior probabilities, compared to their priors, as is the case for (people) in Figure 4. Formally, *selectional* association is defined as:

$$A_R(p, c) = \frac{1}{S_R(p)} * \Pr(c | p) * \log \frac{\Pr(c | p)}{\Pr(c)}$$

See figure 4, we find that the probability distribution ratio will be changed when a new word appears next to a word given previous.

In Table 1, there is a comparison of a chosen word to assign the semantic role label belonging to the class in LLOCE with arguments from the perspective of human.

Table 1. Selectional association for plausible nouns.

Verb	Noun	A <sub>R</sub> (verb, noun)	Semantic classes
go	road	5.78	M126
test	sugar	3.33	E52
check	line	2.67	J41
run	way	2.38	M125

Table 1 presents a selected sample of Resnik's (1993a) comparison with argument plausibility judgments made by human subjects. What is most interesting here is the way in which strongly selecting verbs "choose" the sense of their arguments? For example, *road* has 3 senses in LLOCE, and belongs to 18 classes in all. In order to approximate its plausibility as the object of *go*, the *selectional* association with *go* was computed for all 18 classes, and the highest value returned in this case (M126). Since only one sense of *road* has this class as an ancestor, this method of determining argument plausibility has, in essence, performed sense disambiguation as a side effect. This observation suggests the following simple algorithm for disambiguation by *selectional* preference. Let *n* be a noun that stands in relationship *R* to predicate *p*, and let {*s*<sub>1</sub>, *s*<sub>*k*</sub>} be its possible senses. For *i* from 1 to *k*, compute:

$$C_i = \{c \mid c \text{ is an ancestor of } s_i\}$$

$$a_i = \max_{c \in C_i} (A_R(p, c))$$

and assign *a*<sub>*i*</sub> as the score for sense *s*<sub>*i*</sub>. The simplest way to use the resulting scores, following Miller et al (George Miller, Martin Chodorow, Shari Landes, Claudia Leacock, and Robert Thomas, 1994), is as follows: if *n* has only one sense, select it; otherwise select the sense *s*<sub>*i*</sub> for which *a*<sub>*i*</sub> is greatest, breaking ties by random choice.

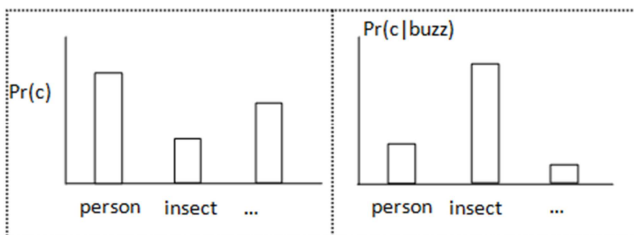


Figure 4. Prior and after distributions over argument of sense.

To illustrate the approach method that we study as follow.

For ability set {*w*<sub>1</sub>, *w*<sub>2</sub> ... *w*<sub>*n*</sub>}, then our algorithm will be built on each pair (*w*<sub>*i*</sub>, *w*<sub>*j*</sub>) with *i* ≠ *j* and algorithm will identify the semantic role for a pair (*w*<sub>*i*</sub>, *w*<sub>*j*</sub>) with the highest semantic similarity. This meaning will be represented by one number that corresponds with quite reasonable meaning of the word. After building all of pairs in the ability set, we will compare each pair whose sense is denoted by a number *x*<sub>*i,k*</sub> for each word *w*<sub>*i*</sub> and that sense is combined with a confidence *c*(*x*<sub>*i,k*</sub>) ∈ [0, 1]. This confidence will be associated with a specific semantic role label. For example in this case, with a bilingual sentence pair “*đường này được làm vào năm 1990*” and “*This road was built in 1990*”, the confidence of pair (*road*, *đường*) will be higher than the confidence of pair (*line*, *đường*). At the end of this step, we highlight the significance of variability in translation: since the method relies on semantic similarities between multiple items in an ability set, the ability set must contain at least two members. Some basis steps are described in figure 5.

**Algorithm 2:**  
 Step 1: Create set *C*.  
 Step 2: For each pair(*v*<sub>*i*</sub>, *S*<sub>*k*</sub>) in *S*  
     Create set *C*<sub>*i*</sub>;  
     For each English word *E*<sub>*n*</sub> in *S*<sub>*k*</sub>  
         Create set *C*<sub>*k*</sub>  
         For each sentence pair in parallel corpus  
             Find word pair (*w*<sub>*v*</sub>, *w*<sub>*e*</sub>) having *w*<sub>*v*</sub> = *v*<sub>*i*</sub>  
             and *w*<sub>*e*</sub> = *E*<sub>*n*</sub>;  
             Adding previous English word of *w*<sub>*e*</sub>  
             into *C*<sub>*k*</sub>;  
             Adding (*v*<sub>*i*</sub>, *E*<sub>*n*</sub>, *C*<sub>*k*</sub>) into *C*;  
             Compute probabilistic ability *x*<sub>*i,k*</sub> and  
             standardized with a confidence *c*(*x*<sub>*i,k*</sub>);  
 Step 3: For each English sentence in parallel  
 corpus  
     For each noun in *S*<sub>*k*</sub>  
         Identify *c*(*x*<sub>*i,k*</sub>) that is highest value  
         Finding correspondence label in LLOCE;  
         Labeling for English noun;

Figure 5. Basis Steps of Algorithm for identifying semantic similarities.

4. Project the sense labels from the target side to the source side of the parallel corpus. we take advantage of the English-side labelling and the word - level alignment to project the semantic labels on English to the corresponding words in Vietnamese. For example, with a bilingual sentence pair “this road was built in 1990” and “*đường này được làm vào năm 1990*”, the result that we obtain is a bilingual sentence pair with the semantic role label such as “this road/<sub>M126</sub> was built in 1990” and “*đường/<sub>M126</sub> này được làm vào năm 1990*”. Label M126 in the semantic label system of LLOCE - LLOCV (Longman Lexicon Of Contemporary English - Longman Lexicon Of Contemporary Vietnamese) will be

presented in the next section.

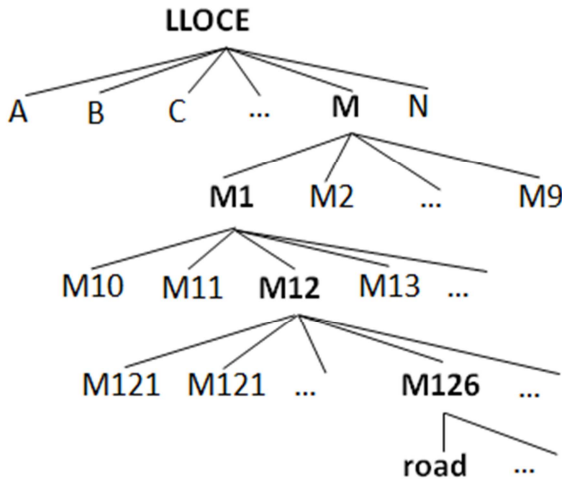


Figure 6. An example of a label in LLOCE.

### 3. Evaluated Approach Method

To set up our approach method, we have relied on the semantic role system in the LLOCE-LLOCV English-Vietnamese bilingual dictionary, which is organized and arranged into 14 themes, each of which is divided into many groups. As a result, there are 129 groups belonging to those 14 themes. Moreover, each group is divided into many classes that include totally 2449 classes (which are also

called semantic classes); and each class is divided into word items - approximately 16000 word items that have related their senses [Dinh Dien, 2006]. Our system will be semantic role labelled for nouns in bilingual English - Vietnamese which belongs to 2449 semantic classes in LLOCE-LLOCV (see figure 6, figure 7).

We use the text mining programs to build corpus semi-automatically on Internet. The texts that we examined have included some fields such as computer science magazine, daily newspaper, token raw data from internet and the other resources quoted from EVC [Dinh Dien and Hoang Kiem, 2003], books (see table 2).

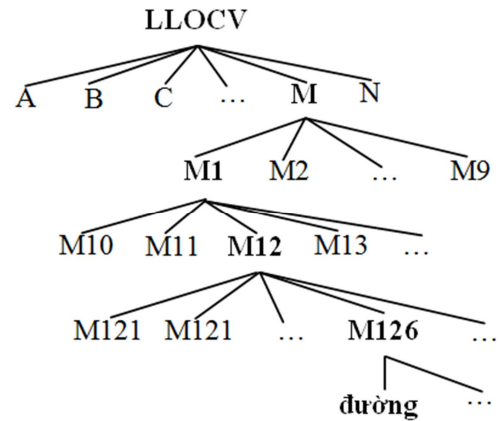


Figure 7. An example of a label in LLOCV.

Table 2. Resources for Experiment.

No.	Resources	The number of pairs of parallel sentences	The number of English words	The number of English nouns
1	Computer Science magazine	1,200	20,902	2,650
2	Token raw data from Internet	1,600	26,870	3,067
3	Daily newspaper	950	10,883	1,879
4	Quote from EVC	3,800	22,800	6,700
5	Other books	1,022	8,472	960
	Total	8,572	89,927	15,256

Data selection criteria for our English and Vietnamese corpus are sentence pairs which are grammatically correct and accepted and used widely. The English-Vietnamese bilingual data must really be 1-1 translations of each other. If they are not 1-1 translations, the computer will be difficult to link the bilingual sentence automatically. Moreover, we should need the 1-1 translations to compare English with Vietnamese words when we project semantic role labels on English to the corresponding words in Vietnamese. However, due to the fact that the amount of data on the internet is huge, the automatic detection of the Webpage containing bilingual English-Vietnamese is not quite easy. Even if there are bilingual websites, determining which pages that are translations of each other is not simple because it requires a lot of resources on languages. However, the resources supporting on Vietnamese are still limited, so we have referred to the

automatic corpus construction methods from Internet on other languages. And then, we have built the data collection method in bilingual English-Vietnamese automatically to cater for this study.

To tackle these problems above, we have used a search engine to quickly find the address of the Web site which contains bilingual English-Vietnamese due to some visual heuristic inspections following:

- Based on the content of links, for example, if an English page links to a Vietnamese page, there will be some words appearing such as “Vietnamese version” or “in Vietnamese” or only “Vietnamese”;
- Based on the hierarchical structure of the page which means the webmaster will be created into a common master page containing links to the sites of other languages;

- Based on the URL of the webpage, bilingual websites often have similar URLs except for the specified language. For example, <http://www.na.gov.vn/htx/English/> and <http://www.na.gov.vn/htx/VietNameese/> is the translated page of each other, or it could also be variables used as [www.hochiminhcity.gov.vn/Pages/default.aspx](http://www.hochiminhcity.gov.vn/Pages/default.aspx) (Vietnamese), [www.eng.hochiminhcity.gov.vn/Pages/default.aspx](http://www.eng.hochiminhcity.gov.vn/Pages/default.aspx) (English).

Then, we have used Google Search Engine to find all the web addresses that contain material signs of both English-Vietnamese languages. After that, the English and Vietnamese text will be downloaded automatically by a Web Crawler program from this site to cater for the next stage of processing.

Corpus that we have chosen should fit the style and the areas in which we are studying, especially the science of technology and the everyday conversations. With the appropriate corpus we will use them to train the processing natural language system in order to meet the best needs of machine translation from English to Vietnamese with the best quality for Vietnamese.

We built the bilingual English-Vietnamese's corpus to training and test system such as: Data in table 2 has been normalized their form (text-only), tone marks (diacritics), character code of Unicode, character font (Times New Roman), etc. Next, this corpus has been sentence aligned and checked spell semi - automatically. An example of our corpus as the following:

N88:5344: Tôi đang đi trên đường một mình.

N88:5344: I am going on the road alone.

Next, we will create ability sets for nouns from this corpus. After that we will measure the semantic similarity to identify the semantic role for nouns. Finally, the system will perform to label the semantic role for English sentences and project them for Vietnamese ones (see figure 8).

To evaluate this approach method, we held-back 210 - sentence part of the training corpus (which has not been used in the training period) with 400 nouns and we achieved the semantic role labels results as follows (see table 3):

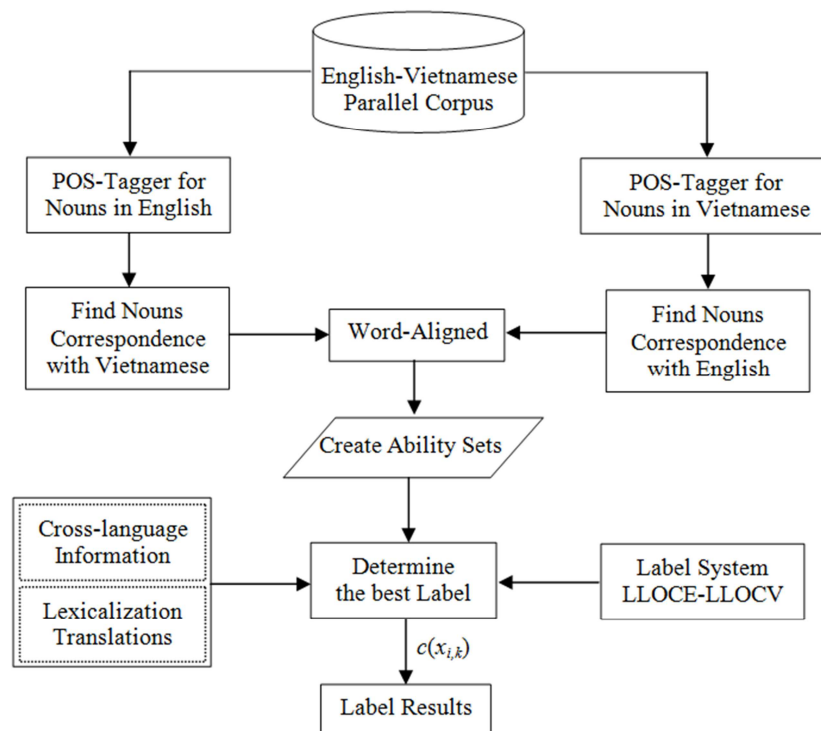


Figure 8. Describe basis components of system.

Table 3. The result of SRLs for experiment.

Correct semantic role labels	incorrect semantic role labels	Precision	Recall
162	28	40.5%	47.5%

Nowadays, there has not been large and standard bilingual corpus yet which were labelled the semantic role on nouns by linguistic experts so that we could use them as a basis in order to evaluate and compare the results on our approach.

Thus, the results of our experiments only describe how to proceed and assign the amount of semantic role labels on the corpus built by statistical machine learning. So the quality of the automatic translation depends on comparing the

similarity of semantic roles (Philip Resnik, 1999) and statistic lexicalization of cross-language transfer (Mikhail Kozhevnikov and Ivan Titov, 2013).

## 4. Discussion

Although the results of our experiments have no corpus to compare and evaluate, the performance of this approach could also be noted. We have built an unsupervised system to semantic role label based on semantic similarity of cross – language which is an important factor in statistic translation, even though those correspondences were derived from machine translations rather than clear human translations. Here we briefly consider issues that bear on recall and precision, respectively.

Some of the sentences in the test corpus could not be automatically aligned because our aligner discards sentence pairs that are longer than a pre-defined limited sentence pairs that are different from the natural language. Moreover, some exceptions for specific signs when translating the language into another language. For these sentences, therefore, no attempt could be made at semantic role label. Our future experiments will attempt to increase the acceptable sentence length, or we will improve our algorithms to separate longer sentences into shorter sentences which will be associated with the special link. When necessary, these sentences can be combined to the complete sentences with their original meanings.

The next issue that we are interest in is building parallel English-Vietnamese corpus, this corpus was semantic role labelled exactly by linguistic experts. When we will use this corpus to evaluate performance of our approach method. Then improving performance of our approach method will be priority in the future research. Another issue that affects the recall is the lack of variability in our method. Of the English nouns that are aligned with source language words, approximately 18% are always aligned with the same word, rendering them unlabelled using an approach based on semantic similarity with target sets.

On inspecting the ability sets qualitatively, we find they contain many outliers, largely owing to noisy alignment. The issue worsens when the outliers are monosemous, since a monosemous word with a misleading sense will erroneously bias the semantic label assignment for the other target set words. These issues reflect the algorithm' implicit assumption that the source words are monosemous, reflected in its attempt to have every word in a ability set influence the semantics of every other word. Inspecting the data produces many counter examples. For example, Vietnamese word {xe đạp} that has the ability set {bicycle, tricycle, bike, motorcycle, velocipede, cyclist}, or word {văn phòng} that

has the ability set {office, living room, meeting, placement}.

## 5. Conclusion and Future Developments

In this paper, we present an unsupervised approach to semantic role label that exploits translations as a proxy for semantic role annotation across language. The observation behind the approach, that words having the same translation often share some dimension of meaning, leads to an algorithm in which the correct meaning of a word is reinforced by the semantic similarity of other words with which it shares those dimension of meaning. In addition, we also exploit the lexicalization translations in cross languages to help identify the semantic role label more appropriately.

Although the content of the article is limited, its contribution has provided an approach for semantic role labelling in bilingual English-Vietnamese. This result supports the automatic machine translation, information retrieval, text summaries etc. Our future researches will effort to improve the performance of the system, especially the accuracy of the semantic role labels. Moreover, we will label the semantic role on verbs, adjectives and adverbs to complete the semantic role labelled system in bilingual English-Vietnamese.

## Acknowledgements

We would like to thank Assoc. Prof. Dr. Dinh Dien (Head of PhD Lab. and Vietnamese Computational Linguistics research Group. Faculty of Information Technology, University of Science, VNU-HCMC) for his guidance as an external advisor on this research, and our colleagues at Ph.D. Lab for the use of their computing facilities in parts of this work.

## References

- [1] Daniel Gildea, Daniel Jurafsky. 2002. Automatic Labelling of Semantic Roles, *2002 Association for Computational Linguistics*.
- [2] Dekang Lin. 2000. Word Sense Disambiguation with a Similarity - Smoothed Case Library, *Computers and the Humanities*, 34: 147-152, 2000.
- [3] Dinh Dien, Hoang Kiem. 2003. POS-Tagger for English-Vietnamese Bilingual Corpus, *Proceedings of the HLT-NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*.
- [4] George Miller, Martin Chodorow, Shari Landes, Claudia Leacock, and Robert Thomas. 1994. Using a semantic concordance for sense identification. In *ARPA Workshop on human Language Technology*, Plainsboro, NJ, March.

- [5] Lluís Marquez, Xavier Carreras, Kenneth C. Litkowski, Suzanne Stevenson. 2008. Semantic Role Labelling: An Introduction to the Special Issue, *2008 Association for Computational Linguistics*, page 145.
- [6] Mikhail Kozhevnikov, Ivan Titov. 2013. Cross-lingual Transfer of Semantic Role Labelling Models, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1190–1200, Sofia, Bulgaria, August 4-9 2013.
- [7] Mona Diab. 2000. An Unsupervised Method for Multilingual Word Sense Tagging Using Parallel Corpora: A Preliminary Investigation. In *SIGLEX2000: Word Sense and Multilinguality*, Hong Kong, October.
- [8] Mona Diab, Philip Resnik. 2002. An Unsupervised Method for Word Sense Tagging using Parallel Corpora, *Proceeding of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, July 2002, pp. 255-262.
- [9] Nancy Ide. 2000. Cross-Lingual Sense Determination: Can It Work? *Computers and the Humanities*, 34: 223-234, 2000.
- [10] Philip Resnik. 1997. Selectional Preference and Sense Disambiguation. In *ANLP Workshop on Tagging Text with Lexical Semantics*, pages 52-54, Washington, D.C., April.
- [11] Philip Resnik. 1999. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language, *Journal of Artificial Intelligence Research 11* (1999) 95-130.