INFORMATION OF DOCTORAL DISSERTATION

PhD student name: LE MANH HAI
Dissertation title: ENGLISH – VIETNAMESE MACHINE TRANSLATION MODEL FOR LEXICAL GAP SOLVING
Major: COMPUTER SCIENCE
Major code: 62.48.01.01
Scientific Advisor: Associate Professor, Doctor. PHAN THI TUOI
School: HCMC University of Technology, Vietnam

DISSERTATION SUMMARY:

A **lexical gap** or **lacuna** is an absence of a word in a particular language. Phenomenon of "Lexical gap" is mentioned in many natural language processing (NLP) studies. For example, English word "abeyant" has meaning "temporarily suspended" in Vietnamese, which has no Vietnamese equivalent single word. In English – Vietnamese machine translation, lexical gap causes many difficulties in terms of both accuracy as well as speed. The thesis proposes an English - Vietnamese model which translates word – to - phrase in case of lexical gap referred as model for English – Vietnamese machine translation in order to solve lexical gap problem. Phrase in the model is a structure with features. When using a structure to convey the information that perhaps only one word, machine translation systems will face enormous obstacles when landing reconstruction with a phrase structure. Information structures from source sentence then are not enough resources to build the target sentence.

Thesis has solved the following problems:
- Study the phenomenon of gaps in vocabulary English-Vietnamese machine translation and the impact of this phenomenon on machine translation
- Building bilingual dictionaries English Vietnamese language support service model to perform a translation from English to a Vietnamese language grammar. This dictionary features a compatible function grammatical structure.
- Propose solutions service model in terms of algorithm processing lexical gap and perform a translation of 200 sentences from English to Vietnamese.

NEW CONTRIBUTIONS / RESULTS OF THE DISSERTATION:

The dissertation proposed a theoretic model and methodology of lexical gap solving for English-Vietnamese machine translation.

* **Result 1**: Building a framework named “Vietnamese lexical functional grammar”, that based on Vietnamese functional grammar. The framework includes new structures with new features.

* **Result 2**: Building a bilingual English – Vietnamese dictionary conforming to Vietnamese lexical functional grammar. This dictionary involves words and phrases in Vietnamese part. Phrases have nested structures.
* **Result 3:** Building a model with processing algorithms for sentence reconstruction, using phrase templates. This model may solve lexical gap problem. The model is main contribution of this thesis

Above results were also presented in published papers of the author.

**APPLICATIONS OF DISSERTATION’S RESULT IN PRACTICE:**

The achieved results of the dissertation lay the foundation of further research of the author. They also support to develop applications in research projects of the author and other machine translation systems and natural language processing studies.

**FURTHER RESEARCH OF THE DISSERTATION:**

There are several matters for further research of the author as follows.

* **Problem 1:** Complete Vietnamese lexical functional grammar as framework.

* **Problem 2:** Lexical gap solving depends on word meaning choice. The latest is not solved yet at the moment

Scientific Advisor                                PhD student

Prof. Dr. PHAN THI TUOI                        LE MANH HAI