

# Determining Semantically Equivalent Questions in a Vietnamese Language Based Document Retrieval System

Dang Tuan NGUYEN, Trung TRAN

Faculty of Computer Science, University of Information Technology, Vietnam National University – HCMC  
Ho Chi Minh City, Vietnam

## Abstract

Generally speaking, in the natural language based document retrieval systems, the ability to determine questions having the equivalent meanings is an important requirement, it allows the optimization of the processing mechanism of questions in related systems. To determine Vietnamese questions having the equivalent meanings for a Vietnamese language based document retrieval system, we have to solve the problems: what is the meaning of a question? And, how can we determine the equivalent meanings of these questions? In this research, we established the method which can allows to determine the Vietnamese questions having the equivalent meanings.

**Keywords:** *Natural Language Processing, Document Retrieval, Syntax, Semantics.*

## 1. Introduction

Generally speaking, in the natural language based document retrieval systems, the ability to determine questions having the equivalent meanings is an important requirement, it allows the optimization of the processing mechanism of questions in related systems. This is also a major challenge for the documents retrieval systems based on Vietnamese Question – Answering model, such as the systems built in [2], [3], [4], [5], [6].

To determine Vietnamese questions having the equivalent meanings for a Vietnamese language based document retrieval system, we have to solve the problems: what is the meaning of a question? And, how can we determine the equivalent meanings of these questions? In this paper, based on rules proposed in “Syntactic Structures” theory of Noam Chomsky [1], we focus on building and testing a method which allows to determine Vietnamese questions having the equivalent meanings, in the domain of document retrieval field.

## 2. Method of Determining Vietnamese Questions having Equivalent Meanings

In this research, we recognize the hypothesis:

- Hypothesis 1: There are Vietnamese questions which their meanings are considered equivalent in the Vietnamese language based document retrieval system.
- Hypothesis 2: The determination of Vietnamese questions having the equivalent meanings have to be implemented based on proper linguistic theory.

Based on “Syntactic Structures” theory of N. Chomsky [1], we developed a method which determines questions having the equivalent meanings for Vietnamese language based document retrieval system. The method that we propose is presented as follows:

Step 1. Analyze the syntactic structures of Vietnamese questions.

Step 2. Determine the syntactic structure of kernel sentence of each question.

Step 3. If these questions are transformed from the same kernel sentence, then determine the sequences of transformations applied to the kernel sentence for generating the considering questions. If these sequences of transformations belong to one of the alternatives determined in Table 1, then the considering questions are considered that they have the equivalent meanings.

In the proposed method above, we recognize that the meaning of a question is regulated by its deep structure. The deep structure of a question is generated by the phrase structure part. In Vietnamese, we do not have morphophonemic rules, therefore the deep structure is also

the structure of kernel sentence which is transformed from the structure of terminal string.

Following Chomsky’s opinion in “Syntactic Structures” [1], the transformations applied for a kernel sentence do not change the meaning of the sentence. However, in this research, we recognize the additional hypothesis: there are transformations considered equivalent when they are applied to kernel sentences having proper structures.

In Table 1, we present the sequences of equivalent transformations in the sense of if these sequences of equivalent transformations have been applied to the same kernel sentence having the determined syntactic structure, then the generated sentences are considered having the equivalent meanings.

Table 1: The sequences of equivalent transformations

	Structure of kernel sentence	Sequences of equivalent transformations
Group 1	danh_ngữ_một + động_từ + danh_ngữ_hai	Tq Tq → Tpass
		Tq → Tw danh ngữ một Tq → Tw danh ngữ một → Tpass
		Tq → Tw danh ngữ hai Tq → Tw danh ngữ hai → Tpass
Group 2	danh_ngữ_một + động_từ + danh_ngữ_hai + bổ_ngữ	Tq Tq → Tpass
		Tq Tq → Tobsep → Tpass
		Tq → Tw danh ngữ một Tq → Tw danh ngữ một → Tpass
		Tq → Tw danh ngữ một Tq → Tw danh ngữ một → Tobsep → Tpass
		Tq → Tw danh ngữ hai Tq → Tw danh ngữ hai → Tpass
		Tq → Tw danh ngữ hai Tq → Tw danh ngữ hai → Tobsep → Tpass
		Tq Tq → Tnominalize
Group 3	danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai +	Tq Tq → Tnominalize

	động_từ_xác_định + danh_từ_tên_riêng  danh_ngữ_một + định_ngữ + động_từ_sở_hữu + danh_ngữ_hai + động_từ_xác_định + danh_từ_tên_riêng	Tq → Tw danh ngữ hai Tq → Tw danh ngữ hai → Tnominalize
Group 4	danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai  danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai + định_ngữ	Tq Tq → Tnominalize
		Tq → Tw danh ngữ một Tq → Tw danh ngữ một → Tnominalize
		Tq → Tw danh ngữ hai Tq → Tw danh ngữ hai → Tnominalize

Transformations Tpass, Tq, Tw, Tobsep, Tnominalize are suitable defined for the application field of theme.

• Transformation Tpass:

Consider the terminal string:

$$NP1 - V - NP2$$

The transformation Tpass will convert this string into following strings:

$$NP2 - được - NP1 - V$$

Or,

$$NP2 - được - V - bởi - NP1$$

For example, consider the active sentence:

Tác giả A viết sách B

Terminal string of this sentence:

Tác giả A – viết – sách B

Apply Tpass, we derive:

Sách B – được – tác giả A – viết  
 Sách B – được – viết – bởi – tác giả A

• Transformation Tq

The transformation Tq will generate the class of Yes/No questions.

Consider the terminal string of affirmative sentence:

Tác giả A – viết – sách B

Apply transformation Tq to this terminal string, we derive the following terminal string:

Tác giả A – viết – sách B

At the result, we derive the interrogative sentence:

Tác giả A viết sách B?

- **Transformation Tw**

Consider the following pair of questions:

Tác giả nào viết tài liệu X  
Tác giả A viết tài liệu nào

To generate these questions, we apply the optional transformation Tw which operate on any string of the form: NP1 – V – NP2.

Transformation Tw can operate in either cases:

Case 1: Tw<sub>1</sub> converts the string of the form “NP1 – V – NP2” into the string of the form “NP1 – nào/gì – V – NP2”.

Case 2: Tw<sub>2</sub> converts the string of the form “NP1 – V – NP2” into the string of the form “NP1 – V – NP2 – nào/gì”.

Transformation Tw can apply only to strings to which Tq has already applied.

Consider the terminal string:

Tác giả A – viết – sách B

Apply transformation Tq to this string, we derive:

Tác giả A – viết – sách B (1)

Then apply transformation Tw to the string (1) in two steps:

Choosing the noun phrase “Tác giả A”, the string (1) will be transformed to (2) by Tw<sub>1</sub>:

Tác giả nào – viết + sách B (2)

Choosing the noun phrase “sách B”, the string (1) will be transformed to (3) by Tw<sub>2</sub>:

Tác giả A + viết – sách nào (3)

- **Transformation Tobsep**

Transformation Tobsep apply to any string of the form:

X – Va – Comp – NP

Which have V → Va + Comp

Consider the sentence:

Tác giả A viết sách B năm Y

In this case, the verb part has the structure “động từ + bổ ngữ” with the verb “viết” and the complement “năm Y”.

To form this sentence, we apply transformation Tobsep to the terminal string:

Tác giả A – viết năm Y – sách B

- **Transformation Tnominalize**

The set of nominalizing transformations will convert sentences into noun phrases. In this paper, we mention the class of possessive sentences.

Example sentence: Tác giả A có sách B

The terminal string of this sentence:

Tác giả A – có – sách B

The transformation Tnominalize converts this string into the following string:

Sách B – của – tác giả A

### 3. Method of determining the sequence of equivalent transformations

In our approach, the sequences of equivalent transformations are defined for each kernel sentence structure. Thus, from one affirmative kernel sentence structure, the sequences of equivalent transformations which are proper defined for this structure will convert it into interrogative sentences having the equivalent meanings.

The generation of interrogative sentences having the equivalent meanings will be illustrated in following samples:

A) Group1:

The structure of kernel sentence:

danh\_ngữ\_một + động\_từ + danh\_ngữ\_hai.

Example 1: Tác giả A viết tài liệu X.

Table 2 show the statistics of sentences transformed from the kernel sentence in example 1.

Table 2: Sentences transformed from the kernel sentence in example 1

<i>Transformational history</i>	<i>Sentences having the equivalent meanings</i>
Tq	Tác giả A viết tài liệu X? Tác giả A đã viết tài liệu X? Tác giả A viết tài liệu X (à / đấy à / nhi / ...)? Tác giả A đã viết tài liệu X (à / đấy à / nhi / ...)? (Có phải là / có đúng là) tác giả A viết tài liệu X (không) (vậy / nhi / ...)? (Có phải là / có đúng là) tác giả A đã viết tài liệu X (không) (vậy / nhi / ...)? Tác giả A viết tài liệu X (có đúng không / có phải không) (vậy / nhi / ...)? Tác giả A đã viết tài liệu X (có đúng không / có phải không) (vậy / nhi / ...)?
Tq → Tpass  (equivalent to the sequence of transformations: Tq)	Tài liệu X được tác giả A viết? Tài liệu X được viết bởi tác giả A? Tài liệu X đã được tác giả A viết? Tài liệu X đã được viết bởi tác giả A? Tài liệu X được tác giả A viết (à / đấy à / nhi / ...)? Tài liệu X được viết bởi tác giả A (à / đấy à / nhi / ...)? Tài liệu X đã được tác giả A viết (à / đấy à / nhi / ...)? Tài liệu X đã được viết bởi tác giả A (à / đấy à / nhi / ...)? (Có phải là / có đúng là) tài liệu X được tác giả A viết (không) (vậy / nhi / ...)? (Có phải là / có đúng là) tài liệu X được viết bởi tác giả A (không) (vậy / nhi / ...)? (Có phải là / có đúng là) tài liệu X đã được tác giả A viết (không) (vậy / nhi / ...)? (Có phải là / có đúng là) tài liệu X đã được viết bởi tác giả A (không) (vậy / nhi / ...)? Tài liệu X được tác giả A viết (có đúng không / có phải không) (vậy / nhi / ...)? Tài liệu X được viết bởi tác giả A (có đúng không / có phải không) (vậy / nhi / ...)? Tài liệu X đã được tác giả A viết (có đúng không / có phải không) (vậy / nhi / ...)? Tài liệu X đã được viết bởi tác giả A (có đúng không / có phải không) (vậy / nhi / ...)?
Tq → Tw danh ngữ một	Ai / tác giả nào / người nào viết tài liệu X Ai / tác giả nào / người nào đã viết tài liệu X? Ai / tác giả nào / người nào viết tài liệu X (vậy / nhi / ...)? Ai / Tác giả nào / Người nào đã viết tài liệu X (vậy / nhi / ...)?

Tq → Tw danh ngữ một → Tpass  (equivalent to the sequence of transformations: Tq → Tw danh ngữ một)	Tài liệu X được ai / tác giả nào / người nào viết? Tài liệu X được viết bởi ai / tác giả nào / người nào? Tài liệu X đã được ai / tác giả nào / người nào viết? Tài liệu X đã được viết bởi ai / tác giả nào / người nào? Tài liệu X được ai / tác giả nào / người nào viết (vậy / nhi / ...)? Tài liệu X được viết bởi ai / tác giả nào / người nào (vậy / nhi / ...)? Tài liệu X đã được ai / tác giả nào / người nào viết (vậy / nhi / ...)? Tài liệu X đã được viết bởi ai / tác giả nào / người nào (vậy / nhi / ...)?
Tq → Tw danh ngữ hai	Tác giả A viết tài liệu nào? Tác giả A đã viết tài liệu nào? Tác giả A viết tài liệu nào (vậy / nhi / ...)? Tác giả A đã viết tài liệu nào (vậy / nhi / ...)?
Tq → Tw danh ngữ hai → Tpass  (equivalent to the sequence of transformations: Tq → Tw danh ngữ hai)	Tài liệu nào được tác giả A viết? Tài liệu nào được viết bởi tác giả A? Tài liệu nào đã được tác giả A viết? Tài liệu nào đã được viết bởi tác giả A? Tài liệu nào được tác giả A viết (vậy / nhi / ...)? Tài liệu nào được viết bởi tác giả A (vậy / nhi / ...)? Tài liệu nào đã được tác giả A viết (vậy / nhi / ...)? Tài liệu nào đã được viết bởi tác giả A (vậy / nhi / ...)?

B) Group 2:

The structure of kernel sentence:

danh\_ngữ\_một + động\_từ + danh\_ngữ\_hai + bổ\_ngữ

Example 2: Tác giả A viết tài liệu X năm Y.

Table 3 show the statistics of sentences transformed from the kernel sentence in example 2.

Table 3: Sentences transformed from the kernel sentence in example 2

<i>Transformational history</i>	<i>Sentences having the equivalent meanings</i>
Tq	Tác giả A viết tài liệu X năm Y? Tác giả A đã viết tài liệu X năm Y? Tác giả A viết tài liệu X năm Y (à / đấy à / nhi / ...)? Tác giả A đã viết tài liệu X năm Y (à / đấy à / nhi / ...)? (Có phải là / có đúng là) tác giả A viết tài liệu X năm Y (không) (vậy / nhi / ...)? (Có phải là / có đúng là) tác giả A đã viết tài liệu X năm Y (không) (vậy / nhi / ...)?

	Tác giả A viết tài liệu X năm Y (có đúng không / có phải không) (vậy / nhi / ...)? Tác giả A đã viết tài liệu X năm Y (có đúng không / có phải không) (vậy / nhi / ...)?
Tq → Tpass  (equivalent to the sequence of transformations: Tq)	Tài liệu X được tác giả A viết năm Y? Tài liệu X được viết bởi tác giả A năm Y? Tài liệu X đã được tác giả A viết năm Y? Tài liệu X đã được viết bởi tác giả A năm Y? Tài liệu X được tác giả A viết năm Y (à / đây à / nhi / ...)? Tài liệu X được viết bởi tác giả A năm Y (à / đây à / nhi / ...)? Tài liệu X đã được tác giả A viết năm Y (à / đây à / nhi / ...)? Tài liệu X đã được viết bởi tác giả A năm Y (à / đây à / nhi / ...)? (Có phải là / có đúng là) tài liệu X được tác giả A viết năm Y (không) (vậy / nhi / ...)? (Có phải là / có đúng là) tài liệu X được viết bởi tác giả A năm Y (không) (vậy / nhi / ...)? (Có phải là / có đúng là) tài liệu X đã được tác giả A viết năm Y (không) (vậy / nhi / ...)? (Có phải là / có đúng là) tài liệu X đã được viết bởi tác giả A năm Y (không) (vậy / nhi / ...)? Tài liệu X được tác giả A viết năm Y (có đúng không / có phải không) (vậy / nhi / ...)? Tài liệu X được viết bởi tác giả A năm Y (có đúng không / có phải không) (vậy / nhi / ...)? Tài liệu X đã được tác giả A viết năm Y (có đúng không / có phải không) (vậy / nhi / ...)? Tài liệu X đã được viết bởi tác giả A năm Y (có đúng không / có phải không) (vậy / nhi / ...)?
Tq → Tobsep → Tpass  (equivalent to the sequence of transformations: Tq)	Tài liệu X được viết năm Y bởi tác giả A? Tài liệu X đã được viết năm Y bởi tác giả A? Tài liệu X được viết năm Y bởi tác giả A (à / đây à / nhi / ...)? Tài liệu X đã được viết năm Y bởi tác giả A (à / đây à / nhi / ...)? (Có phải là / Có đúng là) tài liệu X được viết năm Y bởi tác giả A (không) (vậy / nhi / ...)? (Có phải là / Có đúng là) tài liệu X đã được viết năm Y bởi tác giả A (không) (vậy / nhi / ...)? Tài liệu X được viết năm Y bởi tác giả A (có đúng không / có phải không) (vậy / nhi / ...)? Tài liệu X đã được viết năm Y bởi tác giả A (có đúng không / có phải không) (vậy / nhi / ...)?

	nhi / ...)?
Tq → Tw danh ngữ một	Ai / tác giả nào / người nào viết tài liệu X năm Y? Ai / tác giả nào / người nào đã viết tài liệu X năm Y? Ai / tác giả nào / người nào viết tài liệu X năm Y (vậy / nhi / ...)? Ai / tác giả nào / người nào đã viết tài liệu X năm Y (vậy / nhi / ...)?
Tq → Tw danh ngữ một → Tpass  (equivalent to the sequence of transformations: Tq → Tw danh ngữ một)	Tài liệu X được ai / tác giả nào / người nào viết năm Y? Tài liệu X được viết bởi ai / tác giả nào / người nào năm Y? Tài liệu X đã được ai / tác giả nào / người nào viết năm Y? Tài liệu X đã được viết bởi ai / tác giả nào / người nào năm Y? Tài liệu X được ai / tác giả nào / người nào viết năm Y (vậy / nhi / ...)? Tài liệu X được viết bởi ai / tác giả nào / người nào năm Y (vậy / nhi / ...)? Tài liệu X đã được ai / tác giả nào / người nào viết năm Y (vậy / nhi / ...)? Tài liệu X đã được viết bởi ai / tác giả nào / người nào năm Y (vậy / nhi / ...)?
Tq → Tw danh ngữ một → Tobsep → Tpass  (equivalent to the sequence of transformations: Tq → Tw danh ngữ một)	Tài liệu X được viết năm Y bởi ai / tác giả nào / người nào? Tài liệu X đã được viết năm Y bởi ai / tác giả nào / người nào? Tài liệu X được viết năm Y bởi ai / tác giả nào / người nào (vậy / nhi / ...)? Tài liệu X đã được viết năm Y bởi ai / tác giả nào / người nào (vậy / nhi / ...)?
Tq → Tw danh ngữ hai	Tác giả A viết tài liệu nào năm Y? Tác giả A đã viết tài liệu nào năm Y? Tác giả A viết tài liệu nào năm Y (vậy / nhi / ...)? Tác giả A đã viết tài liệu nào năm Y (vậy / nhi / ...)?
Tq → Tw danh ngữ hai → Tpass  (equivalent to the sequence of transformations: Tq → Tw danh ngữ hai)	Tài liệu nào được tác giả A viết năm Y? Tài liệu nào được viết bởi tác giả A năm Y? Tài liệu nào đã được tác giả A viết năm Y? Tài liệu nào đã được viết bởi tác giả A năm Y? Tài liệu nào được tác giả A viết năm Y (vậy / nhi / ...)? Tài liệu nào được viết bởi tác giả A năm Y (vậy / nhi / ...)? Tài liệu nào đã được tác giả A viết năm Y (vậy / nhi / ...)? Tài liệu nào đã được viết bởi tác giả A năm Y (vậy / nhi / ...)?
Tq → Tw danh ngữ hai → Tobsep → Tpass  (equivalent to the	Tài liệu nào được viết năm Y bởi tác giả A? Tài liệu nào đã được viết năm Y bởi tác giả A? Tài liệu nào được viết năm Y bởi tác giả



sequence of transformations: Tq → Tw danh ngữ hai)	A (vậy / nhi / ...)? Tài liệu nào đã được viết năm Y bởi tác giả A (vậy / nhi / ...)?
--	--

C) Group 3:

The structure of kernel sentence:

danh\_ngữ\_một + động\_từ\_sở\_hữu + danh\_ngữ\_hai + động\_từ\_xác\_định + danh\_từ\_tên\_riêng.

Example 3: Tài liệu X có tác giả là A.

Table 4 show the statistics of sentences transformed from the kernel sentence in example 3.

Table 4: Sentences transformed from the kernel sentence in example 3

<i>Transformational history</i>	<i>Sentences having the equivalent meanings</i>
Tq	Tài liệu X có tác giả là A? Tài liệu X có tác giả là A (à / đấy à / nhi / ...)? (Có phải là / có đúng là) Tài liệu X có tác giả là A (không) (vậy / nhi / ...)? Tài liệu X có tác giả là A (có đúng không / có phải không) (vậy / nhi / ...)?
Tq → Tnominalize (equivalent to the sequence of transformations: Tq)	Tác giả của tài liệu X là A? Tác giả của tài liệu X là A (à / đấy à / nhi / ...)? (Có phải là / có đúng là) tác giả của tài liệu X là A (không) (vậy / nhi / ...)? Tác giả của tài liệu X là A (có đúng không / có phải không) (vậy / nhi / ...)?
Tq → Tw danh ngữ hai	Tài liệu X có tác giả là ai? Tài liệu X có tác giả là ai (vậy / nhi / ...)?
Tq → Tw danh ngữ hai → Tnominalize (equivalent to the sequence of transformations: Tq → Tw danh ngữ hai)	Tác giả của tài liệu X là ai? Tác giả của tài liệu X là ai (vậy / nhi / ...)?

D) Group 4:

The structure of kernel sentence:

danh\_ngữ\_một + động\_từ\_sở\_hữu + danh\_ngữ\_hai.

Example 4: Tác giả A có tài liệu X.

Table 5 show the statistics of sentences transformed from the kernel sentence in example 4.

Table 5: Sentences transformed from the kernel sentence in example 4

<i>Transformational history</i>	<i>Sentences having the equivalent meanings</i>
Tq	Tác giả A có tài liệu X? Tác giả A có tài liệu X (à / đấy à / nhi / ...)? (Có phải là / Có đúng là) tác giả A có tài liệu X (không) (vậy / nhi / ...)? Tác giả A có tài liệu X (có đúng không / có phải không) (vậy / nhi / ...)?
Tq → Tnominalize (equivalent to the sequence of transformations: Tq)	Tài liệu X của tác giả A? Tài liệu X của tác giả A (à / đấy à / nhi / ...)? (Có phải là / có đúng là) tài liệu X của tác giả A (không) (vậy / nhi / ...)? Tài liệu X của tác giả A (có đúng không / có phải không) (vậy / nhi / ...)?
Tq → Tw danh ngữ một	Tác giả nào có tài liệu X? Tác giả nào có tài liệu X (vậy / nhi / ...)?
Tq → Tw danh ngữ một → Tnominalize (equivalent to the sequence of transformations: Tq → Tw danh ngữ một)	Tài liệu X của tác giả nào? Tài liệu X của tác giả nào (vậy / nhi / ...)?
Tq → Tw danh ngữ hai	Tác giả A có tài liệu nào? Tác giả A có tài liệu nào (vậy / nhi / ...)?
Tq → Tw danh ngữ hai → Tnominalize (equivalent to the sequence of transformations: Tq → Tw danh ngữ hai)	Tài liệu nào của tác giả A? Tài liệu nào của tác giả A (vậy / nhi / ...)?

#### 4. Developments

We defined the transformations Tpass, Tq, Tw, Tobsep, Tnominalize for 7 kernel sentence structures (affirmative) which are presented in Table 6. The sequences of equivalent transformations for these kernel sentence structures are presented in Table 1.

Table 6: List of kernel sentence structures (affirmative)

	<i>The structure of kernel sentence</i>	<i>The kernel sentences</i>
Group 1	danh_ngữ_một + động_từ + danh_ngữ_hai	Tác giả A viết tài liệu X
		Nhà xuất bản P phát hành tài liệu X
	danh_ngữ_một + động_từ + danh_ngữ_hai +	Tác giả A viết tài liệu X do nhà xuất bản P phát hành

	định_ngữ	Nhà xuất bản P phát hành tài liệu X do tác giả A viết
Group 2	danh_ngữ_một + động_từ + danh_ngữ_hai + bổ_ngữ	Tác giả A viết tài liệu X năm Y
		Nhà xuất bản P phát hành tài liệu X năm Y
Group 3	danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai + động_từ_xác_định + danh_từ_tên_riêng	Tài liệu X có tác giả là A
		Tài liệu X có nhà xuất bản là P
	danh_ngữ_một + định_ngữ + động_từ_sở_hữu + danh_ngữ_hai + động_từ_xác_định + danh_từ_tên_riêng	Tài liệu X do nhà xuất bản P phát hành có tác giả là A
		Tài liệu X do tác giả A viết có nhà xuất bản là P
Group 4	danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai	Tác giả A có tài liệu X
		Nhà xuất bản P có tài liệu X
	danh_ngữ_một + động_từ_sở_hữu + danh_ngữ_hai + định_ngữ	Tác giả A có tài liệu X do nhà xuất bản P phát hành
		Nhà xuất bản P có tài liệu X do tác giả A viết

We have developed the system which allows to determine the Vietnamese questions having the equivalent meanings based on the method we proposed. We tested the system with more than 898 pair of questions having the equivalent meanings and 1,000 pair of questions having the different meanings. These pairs of questions are selected from 2,260 questions transformed from 14 kernel sentences which are presented in Table 6.

## 5. Conclusions

In this research, based on transformational rules in “Syntactic Structures” theory of N. Chomsky [1], we established the method which can allows to determine the Vietnamese questions having the equivalent meanings. This method is based on the hypothesis in which we can define the sequences of equivalent transformations for determined kernel sentence structures. Thus, to determine two any Vietnamese questions having the equivalent meanings or not, the system will analyze the syntactic structure of each question to determine the

transformational history and the kernel sentence of each one. If these two questions were transformed from the same kernel sentence with sequences of equivalent transformations, then they are considered equivalent meanings.

The result of our research can be developed and applied to the Vietnamese language based document retrieval system has already established in [2], [3], [4], [5], [6].

## References

- [1] Noam Chomsky, Syntactic Structures, Second Edition, Mouton de Gruyter, 2002.
- [2] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "Semantic Model for Representing Vietnamese Questions in OpenCourseWare Retrieval System", Proceedings of the 2011 3rd International Conference on Machine Learning and Computing (ICMLC 2011), vol. 4, February 26-28, 2011, Singapore, pp. 331-335. ISBN: 978-1-4244-9252-7.
- [3] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "Understanding the Vietnamese Questions in OpenCourseWare Retrieval System", Proceedings of the 2011 3rd International Conference on Machine Learning and Computing (ICMLC 2011), vol. 4, February 26-28, 2011, Singapore, pp. 327-330. ISBN: 978-1-4244-9252-7.
- [4] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "Semantic Representation for Processing a Series of Vietnamese Questions in OpenCourseWare Retrieval System", Proceedings of the 2011 3rd International Conference on Computer Engineering and Applications (ICCEA 2011), vol. 2, July 15-17, 2011, Haikou, China, pp. 575-578. ISBN: 978-981-08-9196-1.
- [5] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "A Semantic Approach to Answer Vietnamese Questions in OpenCourseWare Retrieval System", Proceedings of the 2011 International Conference on Software Technology and Engineering (ICSTE 2011), August 12-14, 2011, Kuala Lumpur, Malaysia, pp. 314-320. ISBN-13: 978-0-7918-5979-7.
- [6] Dang Tuan Nguyen, An Hoai Vo, Phuc Tri Nguyen, "Answering a Series of Vietnamese Questions in Library Retrieval System", The 2011 The 2nd International Conference on Future Information Technology (ICFIT 2011), IPCSIT vol. 13 (2011), September 16-18, 2011, Singapore, pp. 445-449. ISSN: 2010-460X.