Semantic Analysis and Theme Determination of Vietnamese Phrase and Sentence Based on Computational and Inferable Methods

Son The PHAM
Department of Information Science and Engineering
University of Information Technology, Vietnam
National University – Ho Chi Minh City
Ho Chi Minh City, Vietnam
Email: sonpt@uit.edu.vn

Dang Tuan NGUYEN
Faculty of Computer Science
University of Information Technology, Vietnam
National University – Ho Chi Minh City
Ho Chi Minh City, Vietnam
Email: dangnt@uit.edu.vn

Abstract — Following the proposition of Reading Answering System Model (RASM), this paper explore a novel computational and inferable method for computing the semantics of Vietnamese phrase and sentence in Vietnamese Question Answering System Model (VietQASM). VietQASM is a Vietnamese question answering system model, which can answer many types of questions about series of events and questions having many interrogative objects. The VietQASM composes four main elements: i) a set of semantic models for phrases; ii) a set of semantic models for sentences; iii) a set of methods for determining the theme of phrase or sentence; iv) a semantic processing mechanism for sentence analysis.

Keywords - Question Answering System, Reading Answering System, Computational Semantics, Semantic Representation, Semantic Model, Vietnamese.

I. INTRODUCTION

The bases of this work concern our computational methods proposed to analyze the semantics of Vietnamese phrase and sentence in [1], [2], [3], [4], [5]. However, in those works, the methods did not process the semantics of phrase. To represent the semantics of phrase, we used semantic (constant) symbols. For this reason, in this paper we aim to propose a novel method to compute the semantics of Vietnamese phrase and sentence. In fact, this method is used to build our Vietnamese Question Answering System Model (VietQASM), which can answer questions about series of events and questions having many interrogative objects.

Therefore, this paper aims to explore a novel method in order to resolve two main issues: i) defining the semantic models for represent the meaning of phrase and sentence; and ii) proposing a mechanism to analyze the semantics of sentence. To confront these challenges, we apply the computational approaches and methods proposed by Patrick Blackburn and Johan Bos [7], [8], [9].

For illustrating, we show some examples, which are based on article titles of VnExpress (http://vnexpress.net) [18] and Tuổi Trẻ Online (http://tuoitre.vn) [19].

II. RELATED WORKS

Patrick Blackburn and Johan Bos [7], [8], [9] introduced some approaches, methods, and techniques to compute and represent the computational and inferable semantics of basic phrases and sentences (statements, questions) in English. In the present, their approaches to compute the semantics do not have any theoretical and experimental solution to process complex structure phrases in English.

Fernando C. N. Pereira and Stuart M. Shieber [10] introduced a method to process Context-Free Grammars (CFG) in Prolog. They used Definite Clause Grammars (DCG) to compute the formal semantics of English phrases and sentences.

Son The Pham and Dang Tuan Nguyen [1], [2] introduced two important elements in VNewsQA/ICT system, which are semantic models and semantic processing mechanism [1] that support the VNewsQA/ICT system to analyze the semantics of Vietnamese sentences. Next, two authors also introduced a Reading Answering System Model (RASM) for Vietnamese language [3] that can read simple sentences to answer relative questions. Implementation of this system was based on a proposed approach of computational semantics [1]. And then, two authors introduced implementation techniques for computing the semantics of Vietnamese sentence in Reading Answering System Model (RASM) [5]. These techniques help for building the syntactic and semantic processing mechanisms of sentences in RASM, which is an innovative model from the traditional Question Answering System model.
III. COMPUTING THE SEMANTICS OF PHRASE

This section introduces some semantic models for computing the semantics of phrases in Vietnamese, based on Reading Answering System Model (RASM) [3], [4], [5], Discourse Representation Theory (DRT) [8], and computational methods of formal semantics [7]. In this paper, we focus on considering following phrases in Vietnamese: Noun Phrase (NP), Quantify Phrase (QuaP), Adjective Phrase (AdjP), Time Phrase (TimeP), Place Phrase (PlaceP), and Verb Phrase (VP).

A. Computing the semantics of noun phrase (NP)

For computing the semantics of noun phrase, we apply the methods of Discourse Representation Theory (DRT) [8] and computational semantics [7] to define a semantic models of some types of simple noun phrase (having no more than five lexicons). The following types of noun phrases are processed:

1) Case 1. Noun phrase structure only includes common nouns:

In this type of noun phrase, the semantic form of common noun is represented as follows [7], [8]:

\[
\text{common_noun}(X)
\]

In which:
- common_noun is a common noun in Vietnamese.
- X is a argument which expresses the meaning of common noun (if any).

Example 1: The semantic forms of some noun phrases are represented as in TABLE I.

<table>
<thead>
<tr>
<th>No.</th>
<th>Noun phrase</th>
<th>Semantic representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>công ty cổ phần</td>
<td>công ty(X)</td>
</tr>
<tr>
<td></td>
<td>(service company)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>viễn thông</td>
<td>viễn thông(X)</td>
</tr>
<tr>
<td>3</td>
<td>chi nhánh ngân hàng</td>
<td>chi nhánh(X)</td>
</tr>
<tr>
<td></td>
<td>(bank branch)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>đất nước</td>
<td>đất nước(X)</td>
</tr>
<tr>
<td></td>
<td>(country)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>quốc gia</td>
<td>quốc gia(X)</td>
</tr>
</tbody>
</table>

2) Case 2. Noun phrase structure includes proper name:

The semantics of proper name is defined as follows:

\[
\text{Proper} \quad \text{name} \rightarrow \{ \text{a set of modifiers represents the meaning of proper name} \}
\]

Modifiers of proper name are common nouns, which are used to modify the meaning of proper name. The semantic form of proper name is defined as follows:

\[
\text{Proper} \quad \text{name} = \{ \text{common_noun}_1(X), \text{common_noun}_2(X), \ldots, \text{common_noun}_n(X) \}
\]

In which:
- common_noun_1, common_noun_2, ... are the semantic forms of common nouns in the set of modifiers.
- X is a proper name.

Example 2: The semantic forms of some noun phrases including proper name are represented as in TABLE II.

<table>
<thead>
<tr>
<th>No.</th>
<th>Noun phrase</th>
<th>Set of modifiers of proper name</th>
<th>Semantic representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Viettel</td>
<td>{tập đoàn, công ty, doanh nghiệp}</td>
<td>{tập đoàn(Viettel), công ty(Viettel), doanh nghiệp(Viettel)}</td>
</tr>
<tr>
<td>2</td>
<td>HSBC</td>
<td>{ngân hàng, tập đoàn}</td>
<td>{ngân hàng(HSBC), tập đoàn(HSBC)}</td>
</tr>
<tr>
<td>3</td>
<td>Việt Nam</td>
<td>{đất nước, quốc gia}</td>
<td>{đất nước(Việt Nam), quốc gia(Việt Nam)}</td>
</tr>
</tbody>
</table>

3) Case 3. Noun phrase structure includes both common nouns and proper names:

This type of noun phrase combines noun phrases of Case 1 and Case 2.

Example 3: The noun phrases “công ty cổ phần FPT” (English: “the FPT joint stock company”) and “ngân hàng thương mại cổ phần ACB” (English: “the Asia Commercial joint stock Bank”) are represented as in TABLE III.

<table>
<thead>
<tr>
<th>No.</th>
<th>Noun phrase</th>
<th>Set of modifier for proper name</th>
<th>Semantic representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>công ty cổ phần</td>
<td>{tập đoàn, công ty, doanh nghiệp}</td>
<td>{tập đoàn(FPT), công ty(FPT), doanh nghiệp(FPT)}</td>
</tr>
<tr>
<td></td>
<td>(FPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ngân hàng thương mại cổ phần ACB</td>
<td>{ngân hàng, tập đoàn, công ty}</td>
<td>{ngân hàng(ACB), tập đoàn(ACB), công ty(ACB)}</td>
</tr>
</tbody>
</table>

4) Case 4. Noun phrase structure includes common nouns, proper names, adjectives or adjective phrases:

This is type of complex noun phrase will be discussed in section C.

5) Case 5. Noun phrase structure based on possessive relationship

This type of noun phrase has a structure as follows:

\[
\text{The first noun phrase} \rightarrow \text{The word } \text{"cúa"} \rightarrow \text{The second noun phrase}
\]

Figure 1. Noun phrase structure based on possessive relationship

In Figure 1., the second noun phrase modifies the first noun phrase. We propose a model to represent the semantics form of this type of noun phrase as follows:

\[
\text{SemNP(<Poss_cúa(<SemNP1>),<SemNP2>}>)
\]

In which:
- SemNP1: the semantics of the first noun phrase.
- SemNP2: the semantics of the second noun phrase.

Example 4: The noun phrase “dich vu cua nga hang Dong A” (English: “service of Dong A bank”).

DOI 10.5013/IJSSST.a.17.32.27
ISSN: 1473-804x online, 1473-8031 print
In which:
- “dịch vụ” is the first noun phrase, and have the semantic form: dịch vụ(X).
- “ngân hàng Đông Á” is the second noun phrase, and have the semantic form: ngân hàng(Dong Á), tập đoàn(Dong Á).

The semantic form of this noun phrase is represented as follows:

\[\text{SemNP}(\text{dịch vụ(X)), ngân hàng(Dong Á), tập đoàn(Dong Á)}\]

6) General structure of noun phrase

The general structure of noun phrase is described as in Figure 2.

![Figure 2. Basic semantic dependencies between constituents of Vietnamese noun phrase](image)

In Figure 2., we have four modifier arrows (A), (B), (C), (D):
- The modifier arrow (A): the adjective phrase has its semantic model as presented in section C.
- The modifier arrow (B): the proper name has its semantic model like of noun phrase in Case 2.
- The modifier arrows (C) and (D): the semantic model is like of noun phrase in Case 1.

With respect to the general noun phrase, its semantic model is a set of the semantic models of Case 1, Case 2, and Case 4.

B. Computing the semantics of quantifier phrase (QuaP)

In Vietnamese, quantifier phrase (QuaP) determine the finiteness or infiniteness of noun phrase. We limit the scope of quantifiers as follows:

- Finite quantifier phrase is number or numerals.
  - Example 5: “một”, “hai”, “ba”, …; 1, 2, 3, etc.

To represent the semantics of quantifier phrase, we define a model to express the meaning of quantifier for noun phrase as follows:

\[\text{ModifyQNP(<SemNP>, <Value of quantifier>)}\]

In which:
- SemNP is semantics of the noun phrase.
- Value of quantifier: If the quantifier is finite, its value is a number. If the quantifier is infinite, its value is “infinite”.

Example 6: “100 nhân viên”, “hầu hết nhân viên” (English: “100 employees”, “a number of employee”). The semantic model of these quantifier phrases are represented as follows:

\[\text{ModifyQNP(<nhân viên(X)>, <100>)}\]

In addition, we also analyze quantifier phrase constructed by noun phrase and ordinal number.

Example 7: “thứ nhất”, “thứ hai”, “lần thứ 8”, … (English: “first”, “second”, “eighth”, …). The semantic models of these quantifier phrases have the form as follows:

\[\text{Modify_Ordinal_NP(<SemNP>, <Value Ordinal>)}\]

In which:
- SemNP is semantics of noun phrase.
- Value ordinal is the value of ordinal number.

Example 8: “Hội nghị ASEAN lần thứ 26” (English: “26th ASEAN Conference”). The semantic model of this quantifier phrase has the form as follows:

\[\text{Modify_Ordinal_NP(<hội_nghị(ASEAN)>, <26>)}\]

C. Computing the semantics of adjective phrase

In English, an adjective modifies a noun; an adverb of manner modifies a verb. However, in Vietnamese, we classify as follows:
- Adjective phrase modifies a noun, in front of noun phrase.
- Adjective phrase modifies a verb (The same as adverb of manner in English).

In this research, we limit the semantic scope of adjective, only analyze types of adjective as follows:

**TABLE IV. TYPES OF ADJECTIVE**

<table>
<thead>
<tr>
<th>No.</th>
<th>Types of adjective</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phạm chát (quality)</td>
<td>tốt, đẹp, xấu, … (good, nice, bad, …)</td>
</tr>
<tr>
<td>2</td>
<td>Màu sắc (color)</td>
<td>đỏ, cam, vàng, lục, … (red, orange, yellow, green, …)</td>
</tr>
<tr>
<td>3</td>
<td>Kích thước (size)</td>
<td>lớn, nhỏ, to, rộng, hẹp, … (big, small, big, wide, narrow, …)</td>
</tr>
<tr>
<td>4</td>
<td>Hình dáng (shape)</td>
<td>tròn, vuông, méo mó, … (circular, square, contorted)</td>
</tr>
<tr>
<td>5</td>
<td>Âm thăng (sound)</td>
<td>ồn ào, nhộn nhịp, … (noisy, crowded, …)</td>
</tr>
<tr>
<td>6</td>
<td>Hương vị (taste)</td>
<td>chua, chát, mặn, ngọt, … (sour, acrid, salty, sweet, …)</td>
</tr>
<tr>
<td>7</td>
<td>Mùi vị (smell)</td>
<td>thơm, tân, hôi, … (perfumed, stinking, smelly, …)</td>
</tr>
<tr>
<td>8</td>
<td>Mức độ (limit)</td>
<td>chậm, nhanh, lẹ, gân, … (slow, fast, quick, near, …)</td>
</tr>
</tbody>
</table>

In this section, we only analyze structure of adjective phrase as follows:

[rất/ thật/ …] + Adjective

In which, [rất/ thật/ …] is equivalent to “very”, or “too”, or “so” in English.
To represent the semantics of adjective phrase, we propose a semantic model having a form as follows:

\[ \text{SemAdjP}(\text{<Adjective>}, \text{<Type of adjective>}) \]

In which:
- Adjective will receive a value, which is lexicon of adjective in phrase.
- Type of adjective will receive a value of adjective type.

Example 9: The adjective phrase “rất tốt” have a semantic model as follows:

\[ \text{SemAdjP}(\text{<tốt>}, \text{<phẩm chất>}) \]

To continue the Case 4 (“noun phrase structure includes common nouns, proper names, and adjective”), we propose a semantic model as follows:

\[ \text{ModifyAdjPNP}(\text{<SemNP>}, \text{<SemAdjP>}) \]

2) Case 2: Structure of time phrase is quantitative time

<table>
<thead>
<tr>
<th>Adjunct</th>
<th>Number</th>
<th>Unit of time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ngày (day)</td>
<td></td>
<td>From 1 to 31</td>
</tr>
<tr>
<td>Tháng (month)</td>
<td></td>
<td>From 1 to 12</td>
</tr>
<tr>
<td>Năm (year)</td>
<td></td>
<td>From 1975 to 2050</td>
</tr>
</tbody>
</table>

To represent the semantics of time phrase in Case 2, we propose a semantic model having a form as follows:

\[ \text{SemTimeP}(\text{date}(D), \text{month}(M), \text{year}(Y)) \]

In which:
- D will receive quantity value of day.
- M will receive quantity value of month.
- Y will receive quantity value of year.

Example 12: Most of the time phrase “khoảng 18 năm”, “khoảng 20 ngày” (English: “about eighteen years”, “around twenty days”). The semantic model is as follows:

\[ \text{SemTimeP}(\text{date}(18), \text{month}_1, \text{year}(18)) \]

E. Computing the Semantics of place phrase

In English, adverb of place and preposition of place are two components of place. In Vietnamese, we focus to analyze structures of place phrase as follows:

[“Taí ở / ở tại / …”] + The common noun of place + The proper name of place

In which:
- [“Taí ở / ở tại / …”] is an adjunct. In English, this is preposition of place such as “in, on, at ...”
- The common noun of place expresses the meaning about region, location, position ... Some of Vietnamese lexicons as follows: “địa điểm”, “địa danh”, “địa phương”, “châu”, “khu vực”, “quốc gia”, “miền”, “vùng”, “thành phố”, “thủ đô”, “quận”, “tỉnh”, “phường”, “huyện”, “thị trấn”, “xã”, “thành phố Hồ Chí Minh”...
- The proper name of place expresses the meaning about place name. For examples: “Hà Nội”, “Sài Gòn”, “Langkawi”...

To represent the semantics of place phrase, we propose a semantic model having a form as follows:

\[ \text{SemPlaceP}(\text{<CNplace>, <PNplace>}) \]

In which:
- CNplace is a semantic form of the common noun of place.
- PNplace is a semantic form of the proper name of place.
NOI”, “in My Tho city”…). The semantic model of place phrase is as follows:
SemPlaceP(<thành_phố(_)>),<_>
SemPlaceP(<thủ_đô(Hà_Nội), thành_phố(Hà_Nội)>,<Hà_Nội>)
SemPlaceP(<thành_phố(Mỹ_Tho), tỉnh(Mỹ_Tho)>,<Mỹ_Tho>)

F. Computing the semantics of verb phrase

In this research, we focus to analyze types of verb in verb phrases:
- Verb has two arguments (“Subject” and “Object”).
- Convert active voice verb to passive voice verb, and contrary.
- Verb phrase have two or three verbs, or two consecutive verbs in which one is inside the other.
- Verb is modified by adverb.

With respect to verb in verb phrase, we have a semantic form as follows:

Verb_lexicon(<Subject>, <Object>)

In which:
- Verb_lexicon is a lexicon of verb.
- “Subject” and “Object” are arguments of verb that will receive value of the semantics of noun phrase or quantify phrase.

Example 14: Verbs “mở”, “thành lập”… (English: “open”, “establish”), which have the semantic form such as mở(X, Y), thành_lập(X, Y).

For verb phrases having many verb or two consecutive verbs in which one is inside the other, the semantic form of the posterior verb is the second argument of the preceding verb. The semantic model of verb phrase is as follows:
SemVP(<SemV>, <Properties>)

In which:
- SemV will receive semantic value of verb.
- Properties will receive a value to determine properties of verb phrase such as passive verbs and adverbs modify verb. Form of properties value is as follows:
VP_Pro(<Adverb>,<Passive>)

In which:
- Adverb: will receive value is lexicon of adverb, if the sentence have adverb.
- Passive: will receive value is lexicon of passive, such as “bị” and “được”, if the is passive voice.

Example 15:
- Verb phrases “thường xuyên thành lập chi nhánh”, “mở cơ sở” (English: “often establish branch”, “open establishment”). The semantic model of verb phrase is as follows:
The semantic model of the first verb phrase:
SemVP(<chuẩn_bị(SemVP(<thành_lập(<>, <chi_nhánh(X)>)),<VP_Pro(<thường_xuyên>,<>)>)),<VP_Pro(<chuẩn_bị(<>, <chi_nhánh(X)>)),<VP_Pro(<><>)>))

G. Determine the theme of phrases

In an effort to keep our system that can represent semantics about event series of sentence. We have to determine and define some themes for phrase in sentence exactly.

First of all, we have to use article titles of VnExpress (http://vnexpress.net) [18] and Tuổi Trẻ (http://tuoitre.vn) [19]. After, we classify the article titles as theme. In this research, we only analyse the theme as follows:
Based on TABLE VII. , we propose a model to represent relation between main theme and sub-theme. The main themes model is as follows:
Main_theme(<Name of main theme>,<Sub-theme>)

In which:
- Name of main theme in above model is name of a theme to determine theme of phrase and sentence.
- Sub-theme is name of a theme, which is in the column (3) of TABLE VII.

Secondly, we more have to define the set of lexicon describe theme of noun phrase.

It is rather difficult to build a set of lexicon to describe the sub-themes. We based on experimental researchs to define the set of lexicon as following:
Note:
- Every element in the set of lexicon (column 3 in TABLE VIII. ) of every theme can be the same.
- The spellings “Vietnamese(English)”: translate Vietnamese into English.
Model is used to represent sub-themes in [TABLE VIII. ] have form as follows:
Sub_Theme(<Name of sub-theme>,<Dictionary>)
### TABLE VII. Relation between main themes and sub-themes in VIETQAS

<table>
<thead>
<tr>
<th>No.</th>
<th>Main Themes</th>
<th>Sub-Themes</th>
</tr>
</thead>
</table>
| 1   | Kinh doanh (Business) | Doanh nghiệp (Enterprise)  
                 Ngân hàng (Bank/Banking)  
                 Thương mại điện tử (E-commerce)  
                 Chứng khoán (Stock)  
                 Bất động sản (Real property) |
| 2   | Kinh tế (Economy)  | Thị trường (Market)  
                 Tài chính (Finance)  
                 Doanh nhân (Businessman/Businesswoman)  
                 Bất động sản (Real property)  
                 Du lịch (Travel)  
                 Khoa học (Science)  
                 Công nghệ (Technology)  
                 Viễn thông (Communication)  
                 Bảo mật (Security)  
                 Di động (Mobile) |
| 3   | Khoa học – Công nghệ (Science - Technology) | Doanh nghiệp (Enterprise)  
                 Ngân hàng (Bank/Banking)  
                 Thương mại điện tử (E-commerce)  
                 Chứng khoán (Stock)  
                 Bất động sản (Real property)  
                 Thị trường (Market)  
                 Tài chính (Finance)  
                 Doanh nhân (Businessman/Businesswoman)  
                 Bất động sản (Real property)  
                 Du lịch (Travel)  
                 Khoa học (Science)  
                 Khoa học – Công nghệ (Science - Technology)  
                 Viễn thông (Communication)  
                 Bảo mật (Security)  
                 Di động (Mobile)  
                 Máy tính (Computer) |

### TABLE VIII. Sets of lexicon describe themes of phrase

<table>
<thead>
<tr>
<th>No.</th>
<th>Theme</th>
<th>Set of lexicon describes theme of noun phrase</th>
<th>Semantic representation of lexicon</th>
</tr>
</thead>
</table>
| 1   | Doanh nghiệp (Enterprise) | Doanh nghiệp (Enterprise)  
                 Tập đoàn (Corporation)  
                 Công ty (Company)  etc...  
                 Doanh nghiệp(X)  
                 Tập đoàn(X)  
                 Công ty(X)  etc...  |
| 2   | Ngân hàng (Bank/Banking) | Ngân hàng (Bank)  
                 Lãi suất (Interest rate)  
                 Ti giá (Exchange rate)  
                 Ngoài tệ (Foreign currency)  etc...  
                 Ngân hàng(X)  
                 Lãi suất(X)  
                 Ti giá(X)  
                 Ngoài tệ(X)  etc...  |
| 3   | Thương mại điện tử (E-commerce) | Thương mại điện tử (E-commerce)  
                 Mua sắm (Shopping)  
                 Trực tuyến (Online)  etc...  
                 Thương mại(X)  
                 Mua sắm(X)  
                 Trực tuyến(X) etc...  |
| 4   | Chứng khoán (Stock) | Chứng khoán (Stock)  
                 Cổ phiếu (Share)  
                 Cổ phiếu (Stock)  etc...  
                 Chứng khoán(X)  
                 Cổ phiếu(X)  etc...  |
| 5   | Bất động sản (Real property) | Bất động sản (Real property)  
                 Bất (Land)  
                 Căn hộ (Apartment)  etc...  
                 Bất động sản(X)  
                 Bất(X)  
                 Căn hộ(X)  etc...  |

In which:
- Name of sub-theme is name of a theme, which is in the column (3) of TABLE VII.
- Dictionary in above model is a set of lexicon to describe the information content relating to theme in the column (4) of TABLE VIII.
Example 16: With theme “Khoa học” (Translate to English: science), we have represent model as follow:
Sub_Theme(  
<Khoa học>,  
<Khoa_học(X), Phát_minh(X), Nghiêng_cứu(X)>  
)

With respect to proper noun and proper name, which also describe theme of noun phrase and represented by model as follow:
Sub_Theme(<Name of sub-theme>,<Pr_Name >)

In which:
- Name of sub-theme is name of a theme, which is in the column (3) of TABLE VII.
- Pr_Name is a proper noun and proper name to describe theme.

Example 17: “Viettel”, “Mobifone” and “Gmobile” are proper names that are in “Science - Technology” main theme and “Communication” sub-theme. Representation model is as follow:
Sub_Theme(<Truyền thông>,<Viettel, Mobifone, Gmobile>)

Finally, we have to define a inferable relation to determine relations between main theme and sub-theme. Inferable method is represented as follows:
Main_theme(<Name of main theme>,<Sub-theme>) → Sub_theme(<Sub-theme>,<Dictionary>)

In which:
- Main theme and Sub_theme are theme models of main theme and sub-theme.
- Symbol “→” represent a inferable relation: if a noun phrase belong Sub-theme then the noun phrase also belong main theme. The contrary is NOT.

Example 18: we analyze noun phrase “viện thông Viettel”. For “viện thông” is an element of “The set of lexicon describe theme of noun phrase” in column (3) [TABLE VIII.] of theme “Communication”, and “Viettel” is proper name that also relate to theme “Communication”, the theme of noun phrase “viện thông Viettel” is “Communication”.

Based on the above inferable model, we infer main theme of “viện thông Viettel” is “Science - Technology”.

IV. COMPUTING THE SEMANTICS AND DETERMINING THE THEME OF SENTENCE

In this section, we based on approaches and methods of computational and inferable semantics [7], [8], [9] proposed for English to build a new computational semantic model for questions about series of events and questions having many interrogative objects in Vietnamese.

In addition, we define a model to represent the themes for sentence.

A. Model of Computational Semantics for Vietnamese Sentence

In Vietnamese, a sentence (statement) is constructed by combining a noun phrase or a quantifier phrase, a verb phrase, and some complementary components (e.g. adjective phrase, place phrase, time phrase).

<table>
<thead>
<tr>
<th>Sentence</th>
<th>NP/QuaP</th>
<th>VP</th>
<th>Complementary components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In which:
- “NP/QuaP” is the semantic expression of noun phrase or quantifier phrase.
- “VP” is the semantic expression of verb phrase.
- “V” is the semantic form of verb.
- “AdjP”, “PlaceP”, “TimeP” are the semantic expression of adjective phrase, place phrase, time phrase.

<table>
<thead>
<tr>
<th>Complementary components</th>
<th>NP/QuaP</th>
<th>V</th>
<th>AdjP</th>
<th>PlaceP</th>
<th>TimeP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Vietnamese, the theme of a sentence will describe something that is a semantic information structure to represent meaning for sentence clearly. There’re a lot of different ways to approach a theme of sentence. But, in this research, we approach to expressed by noun phrases in sentence.

Thematic relation between noun phrases and sentence will clearly represent semantics of sentence more than our previous method. This only is the use of semantic model. A common noun or a proper noun also support to determine themes in sentence.

We based on all of the noun phrases in sentence to determine a theme of sentence. So, a sentence will have many difficult themes, but verbs of sentence don’t support to determine themes of sentence. Semantic model will combine with theme determination model of sentence to create a general semantic model as follow:

General_model(<Semantic model>,<Theme model>)

In which:
- Semantic model is a semantic model of sentence after computing the semantics.
- Theme model is theme determination model of sentence.
C. Processing mechanism for sentence in Vietnamese Question Answering System Model (VietQASM)

In this section, we introduce a new semantic processing mechanism based on the above methods of computational and inferable semantics, theme determination model for phrase and sentence. The semantic processing mechanism for analyzing sentence in our VietQASM is introduced in Figure 3.

![Diagram](https://example.com/diagram.png)

Figure 3. The semantic processing mechanism for analysing a sentence

Notes in Figure 3.:
- The arrow → presents the stage processing sequentially.
- The arrow ↓ presents a support processing.

The semantic processing mechanism of VietQASM for analyzing a sentence includes four processing stages as follows:
- **Stage 1**: Read every lexicon in sentence and use “Rule sets for semantic form of lexicon” to determine its part of speech and its semantic form. A lexicon can have many parts of speeches and semantic forms.
- **Stage 2**: Use “Rule sets for phrase” to determine the type of phrases in sentence, compute the semantics of phrases, and determine theme of phrases in sentence. The result of this stage is a semantic expression. The processing can have many semantic expressions for a phrase.
- **Stage 3**: Use “Rule sets for sentence” to compute the semantics of sentence and infer to determine main theme, sub-theme. The result of this stage is semantic expression of sentence, and can have many semantic expressions for a sentence.
- **Stage 4**: Insert the semantic expression of sentence into the database of the Vietnamese Question Answering System. The data update processing is supported by “Update rule sets”.

We use Definite Clause Grammar (DCG) [7], [8], [10], [11] to build the rule sets “Rule sets for semantic form of lexicon”, “Rule sets for phrase”, “Rule sets for sentence”, “Update rule sets”, and execute this DCG rule sets in SWI-Prolog [12], [13], [14].

Moreover, we based on [15], [16], [17] to develop an application Vietnamese Question Answering System on the domain of Business, Economy, Science-Technology. This system will be built to allow online interaction on the website.

V. EXAMPLES OF SEMANTIC PROCESSING STAGE IN VIETQAS

The elements such as the semantic models of the phrases, the semantic models of the sentences, the semantic processing mechanism for sentence that are applied to analyze the semantic of sentence, which make data in the Vietnamese Question Answering System.

We use an example to illustrate detail of the semantic processing stage.

Example 19: “tập đoàn Viettel mở chi nhánh thứ 2 tại Cần Thơ vào năm 2015” (Translate to English: “the corporation Viettel open the second branch in Can Tho in 2015”).

Based on the semantic models of phrases, the semantic models of sentences, the semantic processing mechanism for sentence in [Figure 3. ], we introduce detail of the semantic processing stage and the semantic processing mechanism as follows:

- **Stage 1**: The “Rule sets for semantic form of lexicon” will support to the system read every lexicon of input sentences. Next, the VietQAS parse part of speech and the semantic form of lexicon. Results after parse as follows:

<table>
<thead>
<tr>
<th>No.</th>
<th>Lexicon</th>
<th>Form Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tập đoàn (corporation)</td>
<td>tập_đoàn(X)</td>
</tr>
<tr>
<td>2</td>
<td>Viettel</td>
<td>tập_đoàn(Viettel) công ty(Viettel) viễn thông(Viettel)</td>
</tr>
<tr>
<td>3</td>
<td>mở (open)</td>
<td>mở(X, Y)</td>
</tr>
<tr>
<td>4</td>
<td>chi nhánh (branch)</td>
<td>chi_nánh(X)</td>
</tr>
<tr>
<td>5</td>
<td>thứ 2 (second)</td>
<td>thứ 2</td>
</tr>
<tr>
<td>6</td>
<td>Cần Thơ</td>
<td>thành phố(Cần_Thơ) tỉnh(Cần_Thơ)</td>
</tr>
<tr>
<td>7</td>
<td>năm (year)</td>
<td>năm(X)</td>
</tr>
<tr>
<td>8</td>
<td>2015</td>
<td>2015</td>
</tr>
</tbody>
</table>

- **Stage 2**: The “Rule sets for phrase” will support for the system read every lexicon of input sentences. Next, the VietQAS parse part of speech and the semantic form of lexicon. Results after parse as follows:
TABLE XI. SEMANTIC REPRESENTATION OF PHRASES

<table>
<thead>
<tr>
<th>No.</th>
<th>Phrases</th>
<th>Semantic representation of phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NP: Tập đoàn Viettel</td>
<td>SemNP(&lt;tập_đoàn(Viettel), công_ty(Viettel), viễn_thông(Viettel)&gt; )</td>
</tr>
<tr>
<td>2</td>
<td>QuanP: chi nhánh thứ 2</td>
<td>ModifyQNP(&lt;chi_nhánh(X)&gt;, &lt;2&gt;)</td>
</tr>
<tr>
<td>3</td>
<td>PlaceP: tại thành phố Cần Thơ</td>
<td>SemPlaceP(&lt;thành_phố(Cần_Thơ), tỉnh(Cần_Thơ)&gt;, &lt;Cần_Thơ&gt; )</td>
</tr>
<tr>
<td>4</td>
<td>TimeP: vào năm 2015</td>
<td>SemTimeP(date(_, _, 2015))</td>
</tr>
<tr>
<td>5</td>
<td>VP: mở chi nhánh thứ 2</td>
<td>SemVP(&lt;mở(&gt;, ModifyQNP(&lt;chi_nhánh(X)&gt;, &lt;2&gt;) &gt;), VP_Pro(&lt;&gt;,&lt;&gt;), Acc_comp(&lt;,&gt;), SemPlaceP(&lt;thành_phố(Cần_Thơ), tỉnh(Cần_Thơ)&gt;, &lt;Cần_Thơ&gt;), SemTimeP(date(_, _, 2015)))</td>
</tr>
</tbody>
</table>

Based on noun phrase “tập đoàn Viettel”. The noun phrase have two themes and theme model of phrase as follows:

**Theme Models of Phrase**

- Sub_Theme(<Doanh nghiệp>,<Tập_đoàn(X)>)
- Sub_Theme(<Viễn thông>,<Viettel>)

**Semantic Model of Sentence**

- SemS(<SemVP(<mở(>, ModifyQNP(<chi_nhánh(X)>, <2>) >), VP_Pro(<>,<>), Acc_comp(<,>), SemPlaceP(<thành_phố(Cần_Thơ), tỉnh(Cần_Thơ)>, <Cần_Thơ>), SemTimeP(date(_, _, 2015)))>)

**Theme Models of Sentence**

- Sub_Theme(<Doanh nghiệp>,<Tập_đoàn(X)>)
- Sub_Theme(<Viễn thông>,<Viettel>)

- Main_theme(<Kinh doanh>,<Doanh nghiệp>) → Sub_Theme(<Đoanh nghiệp>,<Tập_đoàn(X)>)
- Main_theme(<Khoa học-Công nghệ>,<Viễn thông>) → Sub_Theme(<Viễn thông>,<Viettel>)

General semantic model is combination of semantic model of sentence and theme model of sentence as follows:

- General_model(<Semantic Model of Sentence>, Theme Model of Sentence>)

- Stage 4: The “Update rule sets” will support for the system update data from the semantic models of stage 3 into Database of the Vietnamese Question Answering System.

Notes: the system can analyze and represent the semantics of the equivalent sentence as follows:

- “Chi nhánh thứ 2 được mở bởi viễn thông Viettel vào năm 2015 tại thành phố Cần Thơ.”
- “Năm 2015 Viettel mở chi nhánh thứ 2 tại Cần Thơ.”
- “Tai thành phố Cần Thơ Viettel mở chi nhánh thứ 2 vào năm 2015.”
- “Tập đoàn viễn thông Viettel mở chi nhánh thứ 2 vào năm 2015 tại Cần Thơ.”
- “Tai thành phố Cần Thơ vào năm 2015 tập đoàn Viettel mở chi nhánh thứ 2.”
- “Năm 2015 tại tỉnh Cần Thơ chi nhánh thứ 2 được mở bởi tập đoàn viễn thông Viettel.”
- Etc.

VI. CONCLUSION AND FUTURE WORKS

In this research, we introduced some novel semantic models for representing many types of Vietnamese phrases and sentences. Based on these semantic models, we proposed a computational and inferable method to represent the semantics of sentence. Also, we built a method to determine theme of phrase and sentence. This method will be used to analyze the semantics of sentences in VietQASM.

In the future works, we will focus on the view of implementation techniques to compute the semantics of Vietnamese phrases and sentences in VietQASM.

ACKNOWLEDGMENT

This research is funded by Vietnam National University - Ho Chi Minh City (VNU-HCM) under grant number C2015-26-04.
REFERENCES


