International Journal of Advanced Science and Technology
Vol.65 (2014), pp.95-112
http://dx.doi.org/10.14257/ijast.2014.65.08

The Solution for Resolving Inter-sentential Anaphoric Pronoun "nó" in Vietnamese Paragraphs Composing 3 to 5 Simple Sentences

Trung Tran[1] and Dang Tuan Nguyen[2]

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

[1]ttrung@nlke-group.net, [2]dangnt@uit.edu.vn

Abstract

This paper presents a solution for resolving inter-sentential anaphoric pronoun "nó" in Vietnamese paragraphs composing 3 to 5 simple sentences. In Vietnamese, "nó" is a special pronoun, can be used to indicate human, animal or non-animate object depending on the content and context of the paragraph (so in English, we should use "he", "she" or "it") . The proposed solution consists of some finding appropriate antecedent strategies based on constraints about grammatical characteristics of pronoun "nó" and lexicons belonging to noun, verb, adjective in Vietnamese, combine with some improvements in the computational model which we built before for implementing these strategies.

Keywords: *anaphora resolution, discourse representation, inter-sentential anaphora*

1. Introduction

Anaphora resolution, especially resolving inter-sentential anaphoric pronouns in discourses, is an important research topic in natural language processing. In [8] we presented finding antecedent strategies for some inter-sentential anaphoric pronouns indicating human object in Vietnamese paragraphs composing 3 to 5 simple sentences, and a computational model which is implemented in Prolog for implementing these strategies. Proposed finding antecedent strategies based on three factors: 1) constraints about grammatical characteristics in Vietnamese of pronouns indicating human object and lexicons belonging to noun, verb, adjective; 2) the focusing phenomenon when using pronouns standing alone or standing with "ta" / "ấy" / "này". We also built a computational model based on improving the model of Covington and Schmitz [4]. We applied framework Graph Unification Logic Programming (GULP) [3] to develop the model. The computational model consists of four main components: 1) the component for analyzing the syntactic structure tree of sentences in the paragraph – combine with describing grammatical characteristics of each node in this tree based on Unification-based Grammar (UBG) [3, 7]; 2) the component describing grammatical characteristics of the lexicon based on UBG [3, 7]; 3) the component building Discourse Representation Structures (DRS) [1, 6]; 4) the component finding the appropriate antecedent for each pronoun with algorithms based on strategies. In these researched [8] we only resolved pronouns: pronouns indicating human object and standing alone ("anh", "cô", "chị", "ông", "bà", "bạn", "em"), and these pronouns when standing with demonstrative adjectives "ta" / "ấy" / "này" (anh *[ấy / ta / này]*, cô *[ấy / ta / này]*, chị *[ấy / ta / này]*, ông *[ấy / ta / này]*, bà *[ấy / ta / này]*, bạn *[ấy / ta / này]*, em *[ấy / ta / này]*).

ISSN: 2005-4238 IJAST

Follow researches in [8], we proposed in [9] some priority heuristics for resolving these pronouns in two situations which are: when considering one pronoun appearing several times in consecutive sentences; and considering many candidate antecedents of one pronoun. We also presented the approach for integrating these priority heuristics with finding antecedent strategies in [8].

With the similar approach, we presented in [10] finding antecedent strategies for pronouns indicating human objects which are mentioned above and pronoun "nó" in four groups of pairs of simple Vietnamese sentences. With each group of pair of simple Vietnamese sentences we have different finding antecedent strategy and algorithm for implementing this strategy. However, in [10], we accepted that pronoun "nó" only indicates thing (animate or non-animate object), and that is the problem that we have to overcome in this research.

In this paper, follow previous researches in [8, 9, 10], we propose a solution for resolving a special pronoun in Vietnamese is pronoun "nó" (English: "he" / "she" / "it" – depending on the content and context of the paragraph). In Vietnamese, depend on the content and context of the paragraph, when using pronoun "nó", the object which is referred to can be human, animate or non-animate object. In this research, considered Vietnamese paragraphs have some new characteristics in compared with paragraphs which were performed in [8] as follow:

- The number of sentences in the paragraph is in the range from 3 to 5 sentences.
- There is pronoun "nó" in the paragraph.
- Pronoun "nó" can appears one time in one sentence or can appears many times in different sentences of one paragraph.
- The candidate antecedent can be human, animate or non-animate object.
- There can be many candidate antecedents of one pronoun "nó", in which the number of candidate antecedent indicating human object can be more than two.

To determine the appropriate antecedent for each pronoun "nó" appearing in Vietnamese paragraphs which satisfies characteristics in [8] and above new characteristics, we propose a new solution, consisting of two main ideas:

- Posing some new finding antecedent strategies only for resolving pronoun "nó". These new strategies based on constraints about grammatical characteristics of noun, verb, adjective in Vietnamese – which were presented in [8], combine with some supplemented characteristics of noun and pronoun "nó".
- Performing some improvements in the computational model [8] for implementing these new strategies.

2. Strategies for Finding the Antecedent of Pronoun "nó" in Vietnamese Paragraphs

In this section, we present strategies for finding the appropriate antecedent for each pronoun "nó" appearing in Vietnamese paragraphs.

As mentioned above, depending on the content and context of the paragraph, pronoun "nó" can indicate human, animal or non-animate object. Therefore, to determine the appropriate antecedent for each pronoun "nó" in one Vietnamese paragraph, we propose some finding antecedent strategies based on grammatical characteristics in Vietnamese

of pronoun "nó" and lexicons belonging to noun, verb, adjective. In Table 1, we present the classification of constraints for applying to some assumed cases when pronoun "nó" appears in the paragraph:

Table 1. Classify the Constraints

Classification	Assumed cases
Constraints about the pragmatic meaning of pronoun "nó" in the context of Vietnamese paragraph	The pragmatic meaning of pronoun "nó" when considering human objects.
	The pragmatic meaning of pronoun "nó" when considering the non-animate objects.
Constraints about the grammatical role of pronoun "nó" in relationship with the main verb in the sentence	Pronoun "nó" takes the subject role of the main verb in the sentence having the syntactic structure "Sentence → Noun phrase + Verb phrase".
	Pronoun "nó" comes with verb "là" (English: "be") in the sentence having the syntactic structure "Sentence → Noun phrase 1 + là (is) + Noun phrase 2".

Strategies which are classified in Table 1 are detail presented as follow:

2.1. Constraints about the pragmatic meaning of pronoun "nó" in the context of the Vietnamese paragraph

In this research, constraints about the pragmatic meaning of pronoun "nó" in the context of one Vietnamese paragraph are proposed based on following characteristics of Vietnamese: depending on the reality context, when using pronoun "nó" to refer to one human or one non-animate object, the referred object must has some concrete syntactic, semantic and pragmatic characteristics.

We propose two constraints for resolving pronoun "nó" for two assumed cases in the reality context of the paragraphs:

- The first assumed case: Pronoun "nó" is used to refer to one human object in the situation that there are many candidate antecedents indicating human object.
- The second assumed case: Pronoun "nó" is used to refer to one non-animate object in the situation that there are many candidate antecedents indicating non-animate object.

Constraints are proposed to determine the appropriate antecedent for pronoun "nó". These two constraints are detailed presented as follow:

2.1.1. The constraint about the pragmatic meaning of pronoun "nó" when considering human objects: In Vietnamese, when mentioning about one human object, pronoun "nó" will not be used if the mentioned person has the high age or mogul, or is referred to with high level of respect.

Consider the paragraph in Example 1, in which there are two candidate antecedents: one person has the high age or mogul, the other person has low age or mogul.

Example 1: "*Bà Lan có con trai. Nó dễ thương.*"
(English: "Mrs Lan had a son. He is cute.")
→ The first object is "bà Lan" (English: "Mrs Lan") is the person having high age and mogul.
→ The second object is "con trai" (English: "son") is the person having low age or mogul.
→ [nó = con trai].

The finding antecedent strategy: When considering candidate antecedents indicating human object, only consider candidate antecedents which are persons having low age or mogul, or are referred to with low level of respect.

2.1.2. The constraint about the pragmatic meaning of pronoun "nó" when considering non-animate objects: When refer to the non-animate object, pronoun "nó" indicates the real thing, can be real felt and clear identified.

Consider the paragraph in Example 2, in which there are two candidate antecedents indicating non-animate objects: one object cannot be real felt, one object can be real felt and clear identified.

Example 2: "*Nam được mười tám tuổi. Anh mua chiếc xe hơi. Anh thích nó.*"
(English: "Nam is eighteen years old. He buys a car. He likes it.")
→ The first non-animate object is "mười tám tuổi" (English: "eighteen years old") cannot be real felt.
→ The second non-animate object is "chiếc xe hơi" (English: "car"), the existence of this object can be real felt and clear identified.
→ [nó = chiếc xe hơi].

The finding antecedent strategy: When considering candidate antecedents which are non-animate objects, only consider real objects which can be real felt and clear identified.

2.2. The constraints about the grammatical role of pronoun "nó" in relationship with main verb in the sentence

In [8, 9, 10] and in this research, we consider the simple sentences having the syntactic structure belonging to one of three following forms:

- Form 1: Sentence → Noun phrase + Verb phrase
- Form 2: Sentence → Noun phrase + Adjective
- Form 3: Sentence → Noun phrase 1 + "là" (is) + Noun phrase 2

When considering sentences having the syntactic structure Form 1 or Form 3 and there are appearances of pronoun "nó", pronoun "nó" can takes the grammatical role in two assumed cases:

- The first case: Pronoun "nó" takes the subject role of verb in the sentence having the syntactic structure Form 1.
- The second case: Pronoun "nó" comes with verb "là" (English: "be") in the sentence having the syntactic structure Form 3.

Assume that pronoun "nó" which is being considered takes the subject role of the main verb in the sentence or comes with verb "là", in Vietnamese there are constraints denote that the object which is referred to here by pronoun "nó" will be human, animal or non-animate object. We specify two constraints for finding the suitable antecedent for each pronoun "nó" in two assumed cases:

2.2.1. The First Case: Pronoun "nó" is the Subject of the Main Vern in the Sentence Having the Syntactic Structure "Sentence → Noun phrase + Verb phrase": In reality, when considering the sentence having the syntactic structure "Sentence → Noun phrase + Verb phrase" in which pronoun "nó" takes the subject role of verb, the referred object have to be human or animal so that can have the action. Therefore, the finding appropriate antecedent focus will orient to the objects which are human or animal objects in the preceding sentences.

Consider the paragraph in Example 3, in which there are two candidate antecedents: one object is human, one object is non-animate object.

Example 3: "*Nam mua máy tính. Nó học toán.*"
(English: "Nam buys a calculator. He learns math.")
→ Pronoun "nó" appears at the second sentence, taking the subject role of verb "học" (English: "learn").
→ The first candidate antecedent is "Nam" appearing at the first sentence, having the characteristic indicating human object.
→ The second candidate antecedent is "máy tính" (English: "calculator") appearing at the first sentence, having the characteristic indicating non-animate object.
→ [nó = Nam].

The finding antecedent strategy: If pronoun "nó" takes the subject role of the main verb in the sentence having the syntactic structure "Sentence → Noun phrase + Verb phrase", only consider candidate antecedents having the grammatical characteristic indicating human or animal in preceding sentences.

2.2.2. The Second Case: Pronoun "nó" Comes with Verb "là" (is) in the Sentence having the Syntactic Structure "Sentence → Noun phrase 1 + là + Noun phrase 2": In reality, when considering the sentence having the syntactic structure "Sentence → Noun phrase 1+ là + Noun phrase 2", in which pronoun "nó" appears at Noun phrase 1 or Noun phrase 2, the referred object often takes the object role of verb in the sentence. Therefore, the finding appropriate antecedent focus will orient to the object taking the object role of the main verb in the preceding sentence.

Consider the paragraph in Example 4:

Example 4: "Nam có con chó. Nó là giống nòi Phú Quốc."
(English: "Nam has a dog. It is Phú Quốc race.")
→ The second sentence has the syntactic structure "Sentence → Noun phrase 1+ là + Noun phrase 2".
→ Pronoun "nó" appears at Noun phrase 1.
→ The first candidate antecedent is "Nam" taking the subject role of verb "có" (English: "have").
→ The second candidate antecedent is "con chó" (English: "dog") taking the object role of verb "có" (English: "have").
→ [nó = con chó].

The finding antecedent strategy: If pronoun "nó" comes with verb "là" (is) in the sentence having the syntactic structure "Sentence → Noun phrase 1 + là + Noun phrase 2", only consider the object taking the object role of the main verb in the preceding sentence.

3. Improving the computational model of anaphoric pronoun processing

In this section, we present some improvements in the computational model [8] for implementing strategies resolving pronoun "nó" in Vietnamese paragraphs.

About the architecture, we still keep four main components of the model, only perform some improvements:

- Improve the component *Describing grammatical characteristics of lexicon* by Unification-Based Grammar [3, 7]:
  - Describe some more grammatical characteristics of noun.
  - Describe the grammatical characteristics of pronoun "nó".
- Improve the component *Finding the antecedent* with the algorithms based on the strategies:
  - Propose algorithms for finding the antecedent for pronoun "nó" which suitable for strategies proposed in Section 2.

These improvements are detailed presented as follow:

3.1. Improving the Component Describing Syntactic and Semantic Characteristics of Lexicon by Unification-Based Grammar

As presented in Section 2, constraints requires to define some more grammatical characteristics of pronoun "nó" and the lexical belonging to three categories which are noun, verb, adjective. Table 2 presents supplemented grammatical characteristics which are required to be more consistent with strategies proposed above.

Table 2. Grammatical Characteristics of Pronoun "nó" and Lexicon Which Are Required in the Constraints

Constraint	Require grammatical characteristics
Constraint II.A.1	Require noun having the characteristic distinguishing the human object having the high age or mogul with the human object having the low age or mogul.
Constraint II.A.2	Require noun having the characteristic distinguishing the real non-animate object (or can be real felt and clear identified) with the non-real non-animate object (or cannot be real felt and clear identified).
Constraint II.B.1	Require pronoun "nó" having the characteristic distinguishing the subject role with object role of verb.
Constraint II.B.2	Require pronoun "nó" having the characteristic coming with verb "là" in the sentence having the syntactic structure "Noun phrase 1 + là + Noun phrase 2".

In Table 3, we synthetize grammatical characteristics of pronoun "nó" and lexicons belonging to noun, verb, adjective which are described in [8] and new characteristics described in Table 2.

Table 3. Synthetize Grammatical Characteristics of Pronoun "nó" and Lexicons of Three Categories which are Noun, Verb, Adjective

Lexical Categories	Grammatical characteristics
Noun	• Define index `index`: The characteristic distinguishes each object. • Define index `flag_species`: The characteristic distinguishes: human, animal, non-animate object. • Define index `flag_age`: The characteristic distinguishes: the human object has the high age or mogul, the human object has the low age or mogul. • Define index `flag_feeling`: The characteristic distinguishes: the real non-animate object or can be real felt and clear identified, the non-real non-animate object or cannot be real felt and clear identified. • Define index `flag_state`: The characteristic distinguishes: the subject role of verb or adjective, the object role of verb. • Define index `flag_position`: The position characteristic in the paragraph.
Verb	• The semantic meaning of verb in the paragraph. • Define index `flag_position`: The position characteristic in the paragraph. • Define index `arg1`: The characteristic indicates the object taking the subject role of verb. • Define index `arg2`: The characteristic indicates the object taking the object role of verb.
Adjective	• The semantic meaning of adjective in the paragraph. • Define index `flag_position`: The position characteristic in the paragraph. • Define index `arg`: The characteristic indicates the object taking the subject role of adjective.
Pronoun "nó"	• Define index `flag_position`: The position characteristic in the paragraph. • Define index `flag_state`: The characteristic distinguishes: the subject role of verb or adjective, the object role of verb. • Define index `flag_come_with_be`: The characteristic coming with verb "là" in the sentence having the syntactic structure "Sentence → Noun phrase 1 + là + Noun phrase 2".

Above grammatical characteristics are implemented in the model [8] as follows:

3.1.1. Describing Grammatical Characteristics of Noun: Grammatical characteristics of noun which are described in Table 3 will be illustrated through common and proper nouns in paragraphs in Example 1 and Example 2:

- Consider proper noun "Bà Lan" (English: "Mrs Lan") showing the person who has the high age.

  ➔ The grammatical characteristics:

  - Define index `index I` which is unique generated for object "Bà Lan". Index `I` is added to list U in structure DRS of the paragraph. Corresponding, define predicate `named(I,[bà,lan])` and add to list Con in structure DRS of the paragraph.
  - Define index `flag_position` taking the value which is the position of the sentence in the paragraph. Corresponding, define predicate `[position(I,FP)` and add to list Con in structure DRS of the paragraph.
  - Define index `flag_state` taking the value which shows the grammatical role in relationship with the main verb in the sentence. Corresponding, define predicate `state(I,FST)` and add to list Con in structure DRS of the paragraph.
  - Define index `flag_species` taking value `[human]` which shows that object "Bà Lan" is the human object. Corresponding, define predicate `species(I,FSP)` and add to list Con in structure DRS of the paragraph.
  - Define index `flag_age` taking value `[old_age]` which shows that object "Bà Lan" is the old person. Corresponding, define predicate `age_gen(I,FSA)` and add to list Con in structure DRS of the paragraph.

➔ These grammatical characteristics which are described based on the model [8] are presented in Figure 1. In compare with [8, 9, 10], in this research we added index `flag_age`. Therefore, the description is similar to in [8, 9, 10] as follows:

```
n(N) --> [bà,lan],
{
  append([position(I,FP), state(I,FST), species(I,FSP),
age_gen(I,FSA), named(I,[bà,lan])], Con,  NewCon),
  unique_integer(I),
  FSP = [human],
  FSA = [old_age],
  N = syn~(index~I ..
           flag_position~FP ..
           flag_state~FST ..
           flag_species~FSP ..
           flag_age~FSA ..
           class~proper) ..
       sem~(in~ DRSList ..
           out~ NewDRSList)
}.
```

Figure 1. Describe characteristics of proper noun "Bà Lan" in Prolog based on GULP [3]

- Consider common noun "mười tám tuổi" (English: "eighteen years old") showing the non-animate object cannot be real felt.

➔ The grammatical characteristics:

- Define index `index` which is unique generated for object "mười tám tuổi". Index `I` is added to list U in structure DRS of the paragraph. Corresponding, define predicate `mười_tám_tuổi(I)` and add to list Con in structure DRS of the paragraph.
- Define index `flag_position` taking the value which is the position of the sentence in the paragraph. Corresponding, define predicate `[position(I,FP)` and add to list Con in structure DRS of the paragraph.
- Define index `flag_state` taking the value which shows the grammatical role in relationship with the main verb in the sentence. Corresponding, define predicate `state(I,FST)` and add to list Con in structure DRS of the paragraph.
- Define index `flag_species` taking value `[unanimate_object]` which shows that object "mười tám tuổi" is the non-animate object. Corresponding, define predicate `species(I,FSP)` and add to list Con in structure DRS of the paragraph.
- Define index `flag_feeling` taking value `[unreal_feeling]` which shows that object "mười tám tuổi" is the non-animate cannot be real felt. Corresponding, define predicate `feeling(I,FSF)` and add to list Con in structure DRS of the paragraph.

➔ These grammatical characteristics which are described based on the model [8] are presented in Figure 2. In compare with [8, 9, 10], in this research we added index `flag_feeling`. Therefore, the description is similar to in [8, 9, 10] as follows:

```
n(N) --> [mười,tám,tuổi],
{
  append([position(I,FP),  state(I,FST),  species(I,FSP),
feeling(I,FSF), mười_tám_tuổi(I)], Con, NewCon),
  unique_integer(I),
  FSP = [unanimate_object],
  FSF = [unreal_feeling],
  N = syn~(index~I ..
           flag_position~FP ..
           flag_state~FST ..
           flag_species~FSP ..
           flag_feeling~FSF ..
           class~common) ..
       sem~(in~ [drs(U,Con)|Super] ..
           out~ [drs([I|U],NewCon)|Super])
}.
```

Figure 2. Describe characteristics of common noun "mười tám tuổi" in Prolog based on GULP [3]

3.1.2. Describing the Characteristic Coming with Verb "là" (is) of Pronoun "nó": With the characteristic coming with verb "là" (English: "is") of pronoun "nó" in the sentence having the syntactic structure "Sentence → Noun phrase 1 + là + Noun phrase 2", we define the index `flag_come_with_be` in analyzing the structure of the sentence into smaller constituents. This index will takes the value depending on the analyzing structure form of the sentence:

- Take value `[come]` in analyzing form "Sentence → Noun phrase 1 + là + Noun phrase 2".

- Take value `[not_come]` in analyzing form "Sentence → Noun phrase + Verb phrase" and "Sentence → Noun phrase + Adjective".
- When describing characteristics of pronoun "nó", add index `flag_come_with_be` and takes the value which is transferred from the sentence.
- When analyzing the verb phrase into verb and noun phrase, add index `flag_come_with_be` and takes the value which is transferred from the sentence.

Analyzing syntactic structure and describing characteristics of each sentence will be detail described as follow:

- Consider the sentence having the syntactic structure form "Sentence → Noun phrase + Verb phrase":

➔ The analyzing syntactic structure which is described based on the model [8] is presented in Figure 3. In compare with [8, 9, 10], in this research we added index `flag_come_with_be` and set index `flag_position` for verb phrase. Therefore, the description is similar to in [8, 9, 10] as follows:

```
s(S,H1,H3) --> {
   NP = sem~A,
   S = sem~A,
   VP = sem~C,
   NP = sem~scope~C,
   NP = syn~index~D,
   VP = syn~arg1~D,
   NP = syn~flag_state~[subject],
   NP = syn~flag_come_with_be~[not_come],
   S = syn~flag_position~FP,
   NP = syn~flag_position~FP,
   VP = syn~flag_position~FP
},
np(NP,H1,H2), vp(VP,H2,H3).
```

Figure 3. Analyze the sentence having the syntactic structure "Sentence → Noun phrase + Verb phrase" in Prolog based on GULP [3]

➔ Explaining this analyzing:

- Index `index` of noun phrase NP takes value D, which is index `index` is unique generated for each object and transferred from noun.
- Index `arg1` of verb phrase VP takes value D, which is index `index` transferred from noun phrase NP. This index will be transferred to smaller constituents when analyzing the structure of verb phrase.
- Index `flag_state` of noun phrase NP takes value `[subject]`. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_com_with_be` of noun phrase NP takes value `[not_come]`. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_position` of sentence S takes value FP, transferred from analyzing syntactic structure of the paragraph into sentences.

- Index `flag_position` of noun phrase NP takes value FP, transferred from index `flag_position` of sentence S. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_position` of verb phrase VP takes value FP, transferred from index `flag_position` of sentence S. This index will be transferred to smaller constituents when analyzing the structure of verb phrase.

- Consider the sentence having the syntactic structure form "Sentence → Noun phrase + Adjective":

➔ The analyzing syntactic structure which is described based on the model [8] is presented in Figure 4. In compare with [8, 9, 10], in this research we added index `flag_come_with_be`. Therefore, the description is similar to in [8, 9, 10] as follows:

```
s(S,H1,H2) --> {
   S = syn~flag_position~FP,
   NP = syn~flag_position~FP,
   NP = syn~flag_state~[subject],
   NP = syn~flag_come_with_be~[not_come],
   S = sem~A,
   NP = sem~A,
   NP = sem~scope~B,
   Adj = sem~B,
   NP = syn~C,
   Adj = syn~C
},
np(NP,H1,H2), adj(Adj).
```

Figure 4. Analyze the sentence having the syntactic structure "Sentence → Noun phrase + Adjective" in Prolog based on GULP [3]

➔ Explaining this analyzing:

- Index `flag_position` of sentence S takes value FP, transferred from analyzing syntactic structure of the paragraph into sentences.
- Index `flag_position` of noun phrase NP takes value FP, transferred from index `flag_position` of sentence S. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_state` of noun phrase NP takes value `[subject]`. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_com_with_be` of noun phrase NP takes value `[not_come]`. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.

- Consider the sentence having the syntactic structure form "Sentence → Noun phrase 1 + là + Noun phrase 2":

➔ The analyzing syntactic structure which is described based on the model [8] is presented in Figure 5. In compare with [8, 9, 10], in this research we added index `flag_come_with_be` and set index `flag_state` for noun phrase 1 and noun phrase 2. Therefore, the description is similar to in [8, 9, 10] as follows:

```
s(S,H1,H3) --> {
   S = syn~flag_position~FP,
   NP1 = syn~flag_position~FP,
```

```
    NP2 = syn~flag_position~FP,
    NP1 = syn~flag_state~[subject],
    NP2 = syn~flag_state~[object],
    NP1 = syn~flag_come_with_be~[come],
    NP2 = syn~flag_come_with_be~[come],
    S = sem~A,
    NP1 = sem~A,
    NP2 = sem~B,
    NP1 = sem~scope~B,
    NP1 = syn~index~I1,
    NP2 = syn~index~I2,
    NP2 = sem~scope~(in~[drs(U,Con)|Super] ..
    out~[drs(U,[(I1=I2)|Con])|Super])
 },
 np(NP1,H1,H2), [là], np(NP2,H2,H3).
```

Figure 5. Analyze the sentence having the syntactic structure "Sentence → Noun phrase 1 + là + Noun phrase 2" in Prolog based on GULP [3]

➔ Explaining this analyzing:

- Index `flag_position` of sentence S takes value FP, transferred from analyzing syntactic structure of the paragraph into sentences.
- Index `flag_position` of noun phrase NP1 takes value FP, transferred from index `flag_position` of sentence S. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_position` of noun phrase NP2 takes value FP, transferred from index `flag_position` of sentence S. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_state` of noun phrase NP1 takes value `[subject]`. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_state` of noun phrase NP2 takes value `[object]`. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_com_with_be` of noun phrase NP1 takes value `[come]`. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `flag_com_with_be` of noun phrase NP2 takes value `[come]`. This index will be transferred to smaller constituents when analyzing the structure of noun phrase.
- Index `index` of noun phrase NP1 takes value I1, which is index `index` is unique generated for each object and transferred from noun.
- Index `index` of noun phrase NP2 takes value I2, which is index `index` is unique generated for each object and transferred from noun.

- Consider pronoun “nó”:

➔ In Prolog, analyzing the grammatical characteristics of pronoun “nó” based on the model [8]:

```
np(NP,H,H) --> ([nó]),
{
   NP = sem~in~DrsList,
   NP = syn~flag_position~FP,
   NP = syn~flag_state~FST,
   NP = syn~flag_come_with_be~FCB,
   NP = syn~index~Index,
   NP = sem~scope~in~DrsList,
   NP = sem~scope~out~DrsOut,
   NP = sem~out~DrsOut
}.
```

Figure 6. Analyze grammatical characteristics of pronoun “nó” in Prolog based on GULP [3]

➔ Explaining this analyzing:

- Index `flag_position` takes value FP, transferred from analyzing syntactic structure of the paragraph into sentences.
- Index `flag_state` takes value FST, transferred from analyzing syntactic structure of the sentence into smaller constituents.
- Index `flag_come_with_be` takes value FCB, transferred from analyzing syntactic structure of the sentence into smaller constituents.
- Index `index` takes value Index is the unique index of the found antecedent.

3.2. Improving the Component Finding the Antecedent with Algorithms based on Strategies

We propose finding appropriate antecedent algorithms for pronoun “nó” in mentioned cases in Table 1. These proposed algorithms are suitable for strategies presented in Section 2, based on grammatical characteristics of pronoun “nó” and lexicons of noun, verb, adjective, presented in Section 3.1.

These algorithms are presented through pseudo code as follow:

- Consider structure DRS [1, 6] of the paragraph at the time that considering pronoun “nó”, consist of the unique index of each object in list U and the predicates associated with this index in list Con.
- The algorithm suitable for the strategy which is based on Constraint 2.1.1:

```
While (index I is in U)
   While (predicate associated with I is in Con)
      If ((position(I) < position(pronoun)) and
            (species(I) is [human]) and
            (age_gen(I) is [young_age]))
         Index of the antecedent = I
      End If
   End While
End While
```

Figure 7. The finding antecedent algorithm for pronoun "nó" suitable for strategy 2.1.1

➔ Explain the main idea of this algorithm:

- Examine each object having `index` I in list U and the predicates associated with this index I in list Con.
- Check this object satisfying the conditions: has predicate `position(I)` taking the value which lower than `position(pronoun)` of pronoun "nó"; has predicate `species(I)` taking value `[human]` which show that this object is human object; has predicate `age_gen(I)` taking value `[young_age]` which show that this person has low age.
- Set index `index` of the antecedent is I.

- The algorithm suitable for the strategy which is based on Constraint 2.1.2:

```
While (index I is in U)
   While (predicate associated with I is in Con)
      If ((position(I) < position(pronoun)) and
            (species(I) is [unanimate_object]) and
            (feeling(I) is [real_feeling]))
         Index of the antecedent = I
      End If
   End While
End While
```

Figure 8. The finding antecedent algorithm for pronoun "nó" suitable for strategy 2.1.2

➔ Explain the main idea of this algorithm:

- Examine the object having `index` I in list U and the predicates associated with this index I in list Con.
- Check this object satisfying the conditions: has predicate `position(I)` taking the value which lower than `position(pronoun)` of pronoun "nó"; has predicate `species(I)` taking value `[unanimate_object]` which show that this object is non-animate object; has predicate `feeling(I)` taking value `[real_feeling]` which show that this object is real and can be real felt.
- Set index `index` of the antecedent is I.

- The algorithm suitable for the strategy which is based on Constraint 2.2.1:

```
If (flag_state(pronoun) == [subject])
   While (index I is in U)
      While (predicate associated with I is in Con)
         If ((position(I) < position(pronoun)) and
               (species(I) is [human]) and
               (age_gen(I) is [young_age]))
            Index of the antecedent = I
         End If
      End While
   End While
End If
```

Figure 9. The finding antecedent algorithm for pronoun "nó" suitable for strategy 2.2.1

➔ Explain the main idea of this algorithm:

- Examine the object having `index` I in list U and the predicates associated with this index I in list Con.
- Check pronoun "nó" taking the subject role of verb in the sentence: predicate `flag_state(pronoun)` takes value `[subject]`.
- Check this object satisfying the conditions: has predicate `position(I)` taking the value which lower than `position(pronoun)` of pronoun "nó"; has predicate `species(I)` taking value `[human]` which show that this object is human object; has predicate `age_gen(I)` taking value `[young_age]` which show that this object has low age.
- Set index `index` of the antecedent is I.

- The algorithm suitable for the strategy which is based on Constraint 2.2.2:

```
If (flag_come_with_be(pronoun) == [come])
   While (index I is in U)
      While (predicate associated with I is in Con)
         If ((position(I) < position(pronoun)) and
               (state(I) is [object]))
            Index of the antecedent = I
         End If
      End While
   End While
End If
```

Figure 10. The finding antecedent algorithm for pronoun "nó" suitable for strategy 2.2.2

➔ Explain the main idea of this algorithm:

- Examine the object having `index` I in list U and the predicates associated with this index I in list Con.
- Check pronoun "nó" comes with verb "là" in the sentence having the syntactic structure "Sentence → Noun phrase 1 + là + Noun phrase 2": predicate `flag_come_with_be(pronoun)` takes value `[come]`.
- Check this object satisfying the conditions: has predicate `position(I)` taking the value which lower than `position(pronoun)` of pronoun "nó"; has predicate

`state(i)` taking value `[object]` which show that this object takes the object role of verb in the sentence.

- Set index `index` of the antecedent is I.

4. Experiment

We use testing data is 123 Vietnamese paragraphs which are already used for testing in [8, 9]. The system auto generated structure DRS [1, 6] and exactly determined the antecedent for pronoun "nó" in 100 paragraphs. Therefore the successful rate attained in the experiment is 81%, in compared with 70% attained in [8] when not integrate the strategies for resolving pronoun "nó". If we integrate with strategies in [9], then the system determines the antecedent for pronouns indicating human objects and pronoun "nó" in 110 paragraphs, which means the successful rate is 89%.

Analyzing the result, we see that paragraphs which are not successfully performed can be classified into following cases:

- There are many pronoun "nó" appearing in the paragraph. As presented above, depend on the content and context of the paragraph, pronoun "nó" can indicates human, animal or non-animate object. Therefore, to exactly determine the appropriate antecedent for each pronoun "nó" appearing in the paragraph, require to have some more assumptions based on the context of the paragraph.
- There are many candidate antecedents for one pronoun "nó". With above strategies, in many cases, can only reduce the candidate antecedents which do not satisfy the constraints, but there are still others. To determine the appropriate antecedent, require to have some more assumptions based on the context of the paragraph.
- There is no any candidate antecedent appearing precede the pronoun. In these cases, the anaphoric pronouns can appears at the head of first sentence, so there is no candidate antecedent for these pronouns.

5. Discussion

Pronoun "nó" is the special pronoun in Vietnamese, depend on the content and context of the paragraph it can indicates human, animal or non-animate object. To determine the appropriate antecedent for pronoun "nó", we based on some grammatical characteristics of pronoun "nó" and lexicons in the sentence and paragraph. These characteristics are expressed as the constraints in the computational model.

Implement algorithms based on strategies presented in Section 2, the system determines the first found object in the paragraph satisfying the constraint is the appropriate antecedent for considering pronoun "nó". This point leads to the cases which are difficult to exactly determine the appropriate antecedent:

- Pronoun "nó" appears at the different sentences in the paragraph, these sentences can have the consecutive position or not. In reality, the antecedent of these pronoun "nó" can be different, but when applying the proposed strategies, only refer to one antecedent for all these pronoun "nó".

Example 5: "*Lan thích căn nhà. Nó xinh xắn. Cô đi xe đạp. Nó tiện dụng.*"
(English: "Lan likes the house. It is nice. She use the bycicle. It is useful.")
→ The first pronoun "nó" appears at the second sentence.
→ The second pronoun "nó" appears at the forth sentence.
→ Apply the proposed strategies: [nó (1) = căn nhà], [nó (2) = căn nhà].
→ In reality: [nó (1) = căn nhà], [nó (2) = xe đạp].

- There are many candidate antecedents for on pronoun "nó". In reality, the number of candidate antecedents satisfying the constraints can be more than one, these candidates can be human, animal or non-animate object and appear in different sentences. When applying proposed strategies, determine the first found object satisfying constraints as the antecedent for pronoun "nó", this one can be not exactly when consider the content and context of the paragraph.

Example 6: "*Lan sống ở thành phố. Cô về quê. Cô yêu cây cối. Cô mua căn nhà. Nó xinh xắn.*"
(English: "Lan lived in the city. She came back to the village. She loves trees. She buys a house. It is lovely.")
→ Pronoun "nó" appears at the fifth sentence.
→ There are four candidate antecedents satisfying the constraints: objec "thành phố" (city) appears at the first sentence, object "quê" (village) appears at the second sentence, object "cây cối" (tree) appears at the third sentence, object "căn nhà" (house) appears at the forth sentence.
→ Apply the proposed strategies: [nó = thành phố].
→ In reality: [nó = căn nhà].

Therefore, to exactly determine the appropriate antecedent for each pronoun "nó" in Vietnamese paragraph, beside constraints consisting of grammatical characteristics of pronoun "nó" and lexicons belonging to three categories which are noun, verb, adjective, need to propose heuristics with priority depending on the context of the paragraph. These heuristics are proposed when considering pronoun "nó" appearing in different sentences in the paragraph or when considering many candidate antecedents of one pronoun "nó" all satisfying constraints.

6. Conclusion

In this paper, we presented some strategies for resolving pronoun "nó" in Vietnamese paragraphs composing 3 – 5 simple sentences. Proposed strategies consist of constraints based on grammatical characteristics of pronoun "nó" and lexicons belonging to three categories which are noun, verb, adjective in the sentence and paragraph. We also presented some improvement in the model [8] to implement these strategies.

The experiment shows that, applying proposed strategies help for determining the appropriate antecedent for pronoun "nó" in major Vietnamese paragraphs accordant with accordance with the described criteria.

In Discussion, we present some limitations in the current approach and propose the solution for better resolving pronoun "nó" in Vietnamese paragraphs.

In next researches, we will consider cases that difficult to determine the appropriate antecedent of pronoun "nó" mentioned in Discussion.

References

[1] H. Kamp, "A theory of truth and semantic representation", University of Amsterdam: Formal methods in the study of language, **(1981)**, pp. 277-322.
[2] M. Johnson and E. Klein, "Discourse, anaphora and parsing", USA: Center for the Study of Language and Information, Stanford University, Rep. CSLI-86-63, **(1986)** [Also in Proceedings of Coling86 669-675].
[3] M. A. Covington, "GULP 4: An Extension of Prolog for Unification Based Grammar", USA: Artificial Intelligence Center, The University of Georgia, Research Rep. AI-1994-06, **(2007)**.
[4] M. A. Covington and N. Schmitz, "An Implementation of Discourse Representation Theory", USA: Advanced Computational Methods Center, The University of Georgia, **(1989)**.
[5] M. A. Covington, D. Nute, N. Schmitz and D. Goodman, "From English to Prolog via Discourse Representation Theory", USA: The University of Georgia, ACMC Research Rep. 01-0024, **(1988)**.
[6] P. Blackburn and J. Bos, Representation and Inference for Natural Language - Volume II: Working with Discourse Representation Structures, Department of Computational Linguistics, University of Saarland, Germany, **(1999)**
[7] S. M. Shieber, An introduction to unification-based approaches to grammar, Microtome Publishing Brookline, Massachusetts, **(2003)**
[8] T. Tran and D. T. Nguyen, "A computational approach for analyzing inter-sentential anaphoric pronouns in Vietnamese paragraphs", International Journal on Natural Language Computing, 2, 3 **(2013a)**, pp. 23-38, DOI: 10.5121/ijnlc.2013.2303
[9] T. Tran and D. T. Nguyen. Improve Effectiveness Resolving some Inter-sentential Anaphoric Pronouns Indicating Human Objects in Vietnamese Paragraphs using Finding Heuristics with Priority. Proceeding of the 10th IEEE RIVF International Conference on Computing and Communication Technologies (RIVF 2013), **(2013b)**, November 10-13, Ha Noi, Vietnam, pp. 109-114. DOI: 10.1109/RIVF.2013.6719876
[10] T. Tran and D. T. Nguyen. A Solution for Resolving Inter-sentential Anaphoric Pronouns for Vietnamese Paragraphs Composing Two Single Sentences. Proceeding of the 5th IEEE International Conference of Soft Computing and Pattern Recognition (SoCPaR 2013), **(2013c)**, December 15-18, Ha Noi, Vietnam, pp. 172–177.

