

Referent Disambiguation for Anaphoric Human Pronouns in Contexts of Pairs of Vietnamese Sentences Using Affecting and Communicating Verbs

Trung Tran^(✉) and Dang Tuan Nguyen

Faculty of Computer Science, University of Information Technology,
Vietnam National University - Ho Chi Minh City, Ho Chi Minh City, Vietnam
ttrung@nlke-group.net, dangnt@uit.edu.vn

Abstract. In Vietnamese paragraphs, there are two types of personal pronoun which are often used simultaneously in the sentences: type (i) – personal pronouns which stand alone; type (ii) – personal pronouns standing with demonstrative adjectives. This paper considers the contexts of pairs of simple Vietnamese sentences in which there are the simultaneous appearances of two personal pronouns belonging to these above types at the second sentence. The objectives of this research have the following characteristics: the first sentence has one transitive verb having the relationship with two different human objects; the second sentence has one transitive verb having the relationship with two different pronouns, in which the one belonging to type (ii) takes the subject role and the other belonging to type (i) takes the object role. To determine the object in the first sentence which is referred to by each pronoun at the second sentence, we propose a presupposition about the relationship contexts between two verbs. The proposed strategy is based on the idea: the transitive verbs are classified according to two properties which is “affect” and “communication”, then specify the pair of properties of two transitive verbs to determine the antecedent for each pronoun.

Keywords: Anaphoric human pronoun · Ambiguous referent · Verbal relationship context · Discourse representation structure · Semantic representation

1 Introduction

In general, to understand and represent the semantics of a paragraph is an important research section in “abstractive” approach in text summarization field [2, 4, 5, 6, 10, 11, 16]. Especially, the most important part which has to be resolved when doing this study is to determine the exact relationship of a personal pronoun appearing at a sentence and its antecedent appearing at the preceding sentence in some appropriate contexts of the paragraph. To handle this problem, many authors proposed different strategies and methods based on the foundations: Discourse Representation Theory [8, 13, 14, 15, 17], WordNet Ontology [19], Centering Theory [1, 3], and others [7, 9].

Studying the pairs of Vietnamese sentences, which are considered as simplest paragraphs, we initially proposed in [20] the basic strategies and mechanism to perform

these strategies with framework Graph Unification Logic Programming (GULP – [12]) combining between Discourse Representation Theory (DRT – [8, 13, 14, 17]) and Unification-Based Grammar (UBG – [18]). The performing mechanism was proposed with the main phases: (a) analyze the syntactic structure of the paragraph into two separated sentences and set the position for each sentence; (b) analyze the syntactic tree of each sentence and set the appropriate informational characteristics; (c) describe the grammatical characteristics of each lexicon and build the Discourse Representation Structure (DRS) – the central structure of DRT theory which help for representing the relationship between the anaphoric pronoun and its antecedent in the context of each sentence and from this representing the meaning of the paragraph; (d) find the antecedent for each pronoun according to the strategies based on the components of the DRS structure which had been built before.

Based on performing mechanism in [20], we considered in [22, 23] the sentential pairs having the special characteristics with the contexts in which there are the appearances of pronoun “*nó*” – a special pronoun in Vietnamese – and the contexts in which using relative clauses. We established the strategies for resolving the pronouns which were more detail than as in [20] with the priority orders: (i) determine that pronoun “*nó*” refers to animate or non-animated object at the first sentence; (ii) determine pronoun “*nó*” refers to object having smaller age in the identified context; (iii) determine the only pronoun indicating person standing with demonstrative adjective [“*ta*”/“*ấy*”/“*này*”] refers to the human object taking the object role at the first sentence. With these strategies, we improved the performing mechanism: added some characteristics in analyzing the syntactic tree of the sentence; added some grammatical characteristics of lexicon; adjusted the antecedent finding algorithm.

In this research, we consider the following contexts of pairs of sentences: the first sentence has one transitive verb having the relationships with two different persons; the second sentence has one transitive verb having the relationship with two different pronouns indicating person, in which the pronoun belonging to type (ii) takes the subject role and the pronoun belonging to type (i) takes the object role. In this context, we focus on two types of pronouns indicating person: type (i) are pronouns [“*anh*”/“*cô*”/“*chị*”/“*ông*”/“*bà*”/“*bạn*”/“*em*”] standing alone; and type (ii) are these pronouns standing with demonstrative adjective [“*ta*”/“*ấy*”/“*này*”]. We establish a presupposition based on reality experience in using Vietnamese: a transitive verb often has one of two properties “affect” and “communication” in the context of sentence. Then, we classify the considered sentential pairs into four groups according to concrete context:

- Group A: the first transitive verb has property “affect”; the second transitive verb has property “affect”. Example 1: “*Nhân chăm sóc em trai. Anh ta dạy dỗ anh.*” (English: “*Nhân cares for the brother. He teaches him.*”)
- Group B: the first transitive verb has property “affect”; the second transitive verb has property “communication”. Example 2: “*Nhân giúp đỡ Nghĩa. Anh ta cảm ơn anh.*” (English: “*Nhân helps Nghĩa. He thanks him.*”)
- Group C: the first transitive verb has property “communication”; the second transitive verb has property “affect”. Example 3: “*Nhân gặp bác sĩ. Ông ta khám cho anh.*” (English: “*Nhân meets the doctor. He examines him.*”)

- Group D: the first transitive verb has property “communication”; the second transitive verb has property “communication”. Example 4: “*Huấn luyện viên gặp Nhân. Ông ấy hướng dẫn anh.*” (English: “*The coach meets Nhân. He guides him.*”)

With the contexts of these groups, the main idea of strategies for resolving each pronoun is: determine the antecedent is one of two human objects at the first sentence satisfying the appropriate characteristics about the position in the paragraph, relationship role with the first transitive verb, and property of each verb.

To perform these strategies, we propose some improvements in the techniques and algorithm of resolving anaphoric pronouns in [20] with some main points:

- Add the appropriate information in describing characteristics of transitive verb.
- Add the appropriate information in describing characteristics of noun.
- Adjust the characteristic description in analyzing verb phrase.
- Adjust the technique to perform the algorithm for each pair.

The result of the process of determining the antecedent for pronouns at the second sentence is the meaning representation structure of the original pair.

2 The Resolving Process

2.1 The Strategies for Finding the Antecedents

As presented in Section Introduction, in this research, we establish a presupposition based on the reality experience in Vietnamese communication: in the sentence at a concrete context, an action which is expressed by a transitive verb often has one of two properties are “affect” or “communication”. With this presupposition, applying to the considered pairs of sentences, we have the comments:

- The first comment: If two consecutive “affect” or “communication” actions are performed, commonly both actions are done by one object.
- The second comment: If one “affect” action is performed then other “communication” action is performed or vice versa, commonly the pronoun standing alone relates to the object taking the subject role of the first action, the pronoun standing with demonstrative adjective relates to the object taking the object role of the first action. According to above comments, we propose the strategies for resolving pronouns:
- For the context of pairs of sentences belonging to group A

The antecedent finding strategy: because the transitive verb at the first and second sentence both have property “affect” then according to the first comment, in the context of the relationship between two verbs, these verbs are performed by one object.

Therefore we determine the antecedents:

- The pronoun standing with demonstrative adjective taking the subject role refers to the object taking the subject role at the first sentence.
- The pronoun standing alone taking the object role refers to the object taking the object role at the first sentence.

- For the context of pairs of sentences belonging to group B
The antecedent finding strategy: because the transitive verb at the first sentence has property “affect” and the transitive verb at the second sentence has property “communication” then according to the second comment, in the context of the relationship between two verbs, these verbs are performed by different objects. Therefore we determine the antecedents:
 - The pronoun standing with demonstrative adjective taking the subject role refers to the object taking the object role at the first sentence.
 - The pronoun standing alone taking the subject role refers to the object taking the subject role at the first sentence.
- For the context of pairs of sentences belonging to group C
The antecedent finding strategy: because the transitive verb at the first sentence has property “communication” and the transitive verb at the second sentence has property “affect” then according to the second comment, in the context of the relationship between two verbs, these verbs are performed by different objects. Therefore we determine the antecedents:
 - The pronoun standing with demonstrative adjective taking the subject role refers to the object taking the object role at the first sentence.
 - The pronoun standing alone taking the subject role refers to the object taking the subject role at the first sentence.
- For the context of pairs of sentences belonging to group D
The antecedent finding strategy: because the transitive verb at the first and second sentence both have property “communication” then according to the first comment, in the context of the relationship between two verbs, these verbs are performed by one object. Therefore we determine the antecedents:
 - The pronoun standing with demonstrative adjective taking the subject role refers to the object taking the subject role at the first sentence.
 - The pronoun standing with demonstrative adjective taking the subject role refers to the object taking the subject role at the first sentence.

2.2 Improve the Techniques in Mechanism for Implementing the Strategies

To implement the proposed strategies in Sect. 2.1, we apply and propose some improvements in the performing mechanism in [20] as follows:

- In describing grammatical characteristics of transitive verb:
 - Define additional characteristic `flag_property_of_verb`. This characteristic takes value `[affect]` if the verb has property “affect”, take value `[communication]` if the verb has property “communication”.
 - Define additional predicate `property_of_verb`. This predicate will be added to the predicate list of DRS structure. This predicate has two arguments: The first

argument take value [first] or [second] corresponding to the position of the verb at the first or second sentence – this information is transferred down in the syntactic tree when analyzing the paragraph into two separated sentences; the second argument takes value [affect] or [communication] corresponding to the property of the verb (Fig. 1).

```
p(P) --> [khám], {
  append([property_of_verb(FP,FPOV),
    đóng(Arg1,Arg2,CO,CAT,FCLASS,FPOV)],
    Con,NewCon),
  CO = [khám],
  CAT = [action],
  FCLASS = [transitive],
  FPOV = [affect],
  P = syn~(flag_arg1~Arg1 ..
    flag_arg2~Arg2 ..
    flag_position~FP ..
    flag_property_of_verb~FPOV) ..
  sem~(in~ [drs(U,Con)|Super] ..
    out~ [drs(U,NewCon)|Super])
}.
```

Fig. 1. Describe transitive verb “khám” in Example 3 with framework GULP [12]. Characteristic `property_of_verb` takes value [affect]. Predicate `property_of_verb` has two arguments: the first argument takes value FP is the position information will be transferred down in the syntactic tree, the second argument takes value [affect].

- In describing grammatical characteristics of noun:
 - Define additional characteristic `flag_property_of_verb`. The characteristic shows that the noun or pronoun has the relationship with the verb having what property. This characteristic takes value [affect] if the verb has property “affect”, takes value [communication] if the verb has property “communication”.
 - Define additional characteristic `flag_index_other`. We notice that two different pronouns relate to two different objects, we propose a technique in which when determine the antecedent for one pronoun then simultaneously determine the remaining object as the antecedent for the other pronoun. This characteristic takes value as the unique index of the object which is determined as the antecedent of the remaining pronoun. This value is different from the value of characteristic `flag_index` (Fig. 2).
- In analyzing verb phrase into verb and noun phrase:
 - Add the mechanism for transferring value F of characteristic `flag_property_of_verb` of the verb to the noun phrase. This value is transferred down in the syntactic tree to characteristic `flag_property_of_verb` of the noun.
 - Add the mechanism for transferring value G of characteristic `flag_index_other` of the noun phrase to the first argument `flag_arg1` of the transitive verb. The value of characteristic `flag_index` and `flag_index_other` corresponding to the indexes of two antecedents of the second and first pronoun. These two indexes will be determined when implementing the antecedent finding algorithm (Fig. 3).

```

n(N) --> [nhân], {
  append([position(I,FP), role(I,FR), species(I,FCLASS),
        nhân(I,CO,CAT,FCLASS)],
        Con,NewCon),
  unique_integer(I),
  CO = [nhân],
  CAT = [object],
  FCLASS = [human],
  N = syn~(flag_index~I ..
          flag_position~FP ..
          flag_property_of_verb~FPOV ..
          flag_index_other~FIO ..
          flag_role~FR) ..
  sem~(in~ [drs(U,Con)|Super] ..
       out~ [drs([I|U],NewCon)|Super])
}.

```

Fig. 2. Describe characteristics of noun “nhân” in Example 3 with framework GULP [12].

- Propose the algorithm for implementing the antecedent finding strategies. With the notice that when improving describing characteristics of noun, we determine the index of the antecedent for the pronoun standing alone – corresponding to the second pronoun taking the object role of the second transitive verb. This index becomes the value for characteristic `flag_index` of noun. The index of the remaining object will become the value for characteristic `flag_index_other` of noun – corresponding to the antecedent of the remaining pronoun. The general algorithm:

Algorithm 1: Determine the antecedents for pronouns.

Consider DRS structure at current time;

Step 1: Check property of second verb

- Express by value of feature `flag_property_of_verb` of second pronoun.

Step 2: Check property of first verb

- Express by value of predicate `property_of_verb()` of first verb.

Step 3: Check object having index `I` and value of predicate `position(I), species(I), role(I)`

- If group A then these values are `[first], [human], [goal]` respectively.
- If group B then these values are `[first], [human], [agent]` respectively.
- If group C then these values are `[first], [human], [agent]` respectively.
- If group D then these values are `[first], [human], [goal]` respectively.

Step 4: The antecedence of the second pronoun is object having index `I`

- Set feature `flag_index` of second pronoun value `I`;
- Transfer this value to feature `flag_arg2` of the second verb;

Step 5: Check object having index `I_other` and

- `I_other` is different from `I`

Step 6: The antecedence of the first pronoun is object having index `I_other`

- Set feature `flag_index_other` of second pronoun value `I_other`;
 - Transfer this value to feature `flag_arg1` of the second verb;
-

```

pp (PP, H1, H2) --> {
  PP = syn~flag_position~E,
  NP = syn~flag_position~E,
  P = syn~flag_position~E,
  P = syn~D,
  PP = syn~D,
  NP = sem~A,
  PP = sem~A,
  NP = syn~flag_index~C,
  PP = syn~flag_arg2~C,
  NP = syn~flag_index_other~G,
  PP = syn~flag_arg1~G,
  NP = syn~flag_role~[goal],
  PP = syn~flag_property_of_verb~F,
  NP = syn~flag_property_of_verb~F,
  P = sem~B,
  NP = sem~scope~B
},
p(P), np(NP, H1, H2).

```

Fig. 3. Analyze the syntactic of verb phrase with framework GULP [12].

After performing the above improvements, we determine the antecedents for pronouns at the second sentence and complete the DRS structure. Consider the pair of sentences in Example 3, we have the result

- A list contains indexes: index 1 expresses object “nhân”; index 2 expresses object “bác sĩ”.

```

np (NP, H, H) --> ([anh]; [cô]; [chị]; [ông]; [bà]; [bạn]; [em]), {
  NP=sem~in~DrsList,
  NP=syn~flag_property_of_verb~FPOV,
  FPOV == [affect],
  member(drs(U, Con), DrsList),
  member(property_of_verb([first], [affect]), Con),
  member(Index, U),
    member(position(Index2, [first]), Con),
    member(species(Index2, [human]), Con),
    member(role(Index2, [goal]), Con),
    Index == Index2,
  NP=syn~flag_index~Index,
  member(Index_Other, U),
    Index_Other \= Index2,
  NP=syn~flag_index_other~Index_Other,
  NP=sem~scope~in~DrsList,
  NP=sem~scope~out~DrsOut,
  NP=sem~out~DrsOut
}.

```

Fig. 4. Implement Algorithm 1 with framework GULP [12].

- A list contains predicates:
- Express information about the object having index 1:
 - `nhân(1, [nhân], [object], [human])`
 - `species(1, [human])`

- `role(1, [agent])`
- `position(1, [first])`
- Express information about the object having index 2:
 - `bác_sĩ(2, [bác, sĩ], [object], [human])`
 - `species(2, [human])`
 - `role(2, [goal])`
 - `position(2, [first])`
- Express information about the first transitive verb:
 - `gặp(1, 2, [gặp], [action], [transitive], [communication])`
 - `property_of_verb([first], [communication])`
- Express information about the second transitive verb:
 - `khám(2, 1, [khám], [action], [transitive], [affect])`
 - `property_of_verb([second], [affect])`

3 Experiment and Discussions

For testing, we collected 200 pairs of Vietnamese sentences having characteristics which are suitable for the research objective and are classified as: group A has 45 pairs, group B has 47 pairs, group C has 48 pairs, and group D has 60 pairs. The testing results:

- Determine correctly the antecedents and build the DRS structures for 29 pairs of group A. The successful rate is 64.44 %.
- Determine correctly the antecedents and build the DRS structures for 41 pairs of group B. The successful rate is 87.23 %.
- Determine correctly the antecedents and build the DRS structures for 37 pairs of group C. The successful rate is 77.08 %.
- Determine correctly the antecedents and build the DRS structures for 51 pairs of group D. The successful rate is 85 %.

The testing results show that, the system determined the exact antecedents for pairs of pronouns at the major of tested pairs of sentences. Analyzing deeper, we notice some points:

- With sentential pairs of group A, in which there are two “affect” actions are performed consecutively, in some cases the second action is performed by the second object. This situation also happened commonly in reality because the second object will react against the previous action of the first object.
- With sentential pairs of the other groups, because there is lack of some additional factors such as time or space which affect the context, therefore the results maybe not correct.

Besides, we see that the techniques in this research are applied for specific types of pairs of sentences. This requires more improvements so that we can apply for pairs of sentences having more complex characteristics.

4 Conclusion

In this research, we presented strategies and techniques for determining the antecedents for each pairs of pronouns belonging to two types: type (i) are pronouns indicating person standing alone; type (ii) are these pronouns standing with demonstrative adjective. With the classification of pairs of sentences into groups based on two properties “affect” and “communication” of transitive verbs at two sentences, we proposed the appropriate resolving strategies.

The testing results show that the antecedent finding strategies and algorithm are suitable for the major of tested pairs of sentences. We also pointed out some points that need to be improved so that can apply for more complex pairs of sentences.

References

1. Joshi, A.K., Weinstein, S.: “Control of inference: role of some aspects of discourse structure centering”. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 1981), pp. 385–387 (1981)
2. Khan, A., Salim, N.: A Review on Abstractive Summarization Methods. *J. Theor. Appl. Inf. Technol.* **59**(1), 64–72 (2014)
3. Grosz, B.J., Joshi, A.K., Weinstein, S.: Centering: a framework for modeling the local coherence of discourse. *Comput. Linguist.* **21**(2), 203–225 (1995)
4. Saranyamol, C.S., Sindhu, L.: A survey on automatic text summarization. *Int. J. Comput. Sci. Inf. Technol.* **5**(6), 7889–7893 (2014)
5. Das, D., Martins, A.F.T.: A survey on automatic text summarization. Language Technologies Institute, Carnegie Mellon University (2007)
6. Lloret, E.: “Text summarization: an overview”, paper supported by the Spanish Government under the project TEXT-MESS (TIN2006-15265-C06-01) (2008)
7. Cornish, F.: Inter-sentential anaphora and coherence relations in discourse: a perfect match. *Lang. Sci.* **31**(5), 572–592 (2009)
8. Kamp, H.: A theory of truth and semantic representation. In: Jeroen, A.G.G., Janssen, T.M. V., Stokhof, M.B.J. (eds.) *Formal Methods in the Study of Language, Part 1. Mathematical Centre Tracts*, pp. 277–322 (1981)
9. Carbonell, J.G., Brown, R.D.: Anaphora resolution: a multi-strategy approach. In: Proceedings of the 12th International Conference on Computational Linguistics, pp. 96–101 (1988)
10. Jezek, K., Steinberger, J.: Automatic text summarization. In: Snasel, V. (Ed.): *Znalosti 2008*, ISBN 978-80-227-2827-0, FIIT STU Brno, Ustav Informatiky a softveroveho inzinierstva, pp. 1–12 (2008)
11. Jones, K.S.: “Automatic summarising: a review and discussion of the state of the art”, Technical report 679. University of Cambridge, Computer Laboratory (2007)
12. Covington, M.A.: “GULP 4: An Extension of Prolog for Unification Based Grammar”, Research Report number: AI-1994-06. Artificial Intelligence Center, The University of Georgia, USA (2007)
13. Covington, M.A., Schmitz, N.: “An Implementation of Discourse Representation Theory”, ACMC Research Report number: 01-0023. The University of Georgia, Advanced Computational Methods Center (1989)

14. Covington, M.A., Nute, D., Schmitz, N., Goodman, D.: “From English to Prolog via Discourse Representation Theory”, ACMC Research Report number: 01-0024. University of Georgia, Advanced Computational Methods Center (1988)
15. Johnson, M., Klein, E.: “Discourse, anaphora and parsing”, Report number: CSLI-86-63. Center for the Study of Language and Information, Stanford University, USA (1986)
16. Kasture, N.R., Yargal, N., Singh, N.N., Kulkarni, N., Mathur, V.: A survey on methods of abstractive text summarization. *Int. J. Res. Merging Sci. Technol.* **1**(6), 53–57 (2014)
17. Blackburn, P., Bos, J.: Representation and inference for natural language – volume ii: working with discourse representation structures. Department of Computational Linguistics, University of Saarland (1999)
18. Shieber, S.M.: An Introduction to Unification-Based Approaches to Grammar. Microtome Publishing Brookline, Massachusetts (2003)
19. Liang, T., Wu, D.S.: Automatic pronominal anaphora resolution in english texts. *Comput. Linguist. Chin. Lang. Process.* **9**(1), 21–40 (2004)
20. Tran, T., Nguyen, D.T.: A solution for resolving inter-sentential anaphoric pronouns for vietnamese paragraphs composing two single sentences. In: Proceedings of the 5th International Conference of Soft Computing and Pattern Recognition (SoCPaR 2013), Hanoi, Vietnam, pp. 172–177 (2013)
21. Tran, T., Nguyen, D.T.: “Merging two vietnamese sentences related by inter-sentential anaphoric pronouns for summarizing. In: Proceedings of the 1st NAFOSTED Conference on Information and Computer Science, Hanoi, Vietnam, pp. 371–381 (2014)
22. Tran, T., Nguyen, D.T.: Semantic predicative analysis for resolving some cases of ambiguous referents of pronoun “Nó” in summarizing meaning of two vietnamese sentences. In: Proceedings of the UKSim-AMSS 17th International Conference on Computer Modelling and Simulation (UKSim 2015), pp. 340–345. Cambridge, United Kingdom (2015)
23. Tran, T., Nguyen, D. T.: Combined method of analyzing anaphoric pronouns and inter-sentential relationships between transitive verbs for enhancing pairs of sentences summarization. In: Silhavy, R., et al. (eds.) Proceedings of the 4th Computer Science On-line Conference (CSOC2015) – Vol 1: Artificial Intelligence Perspectives and Applications, Advances in Intelligent Systems and Computing – Vol. 347, Faculty of Applied Informatics, Tomas Bata University in Zlin, Czech Republic, pp. 67–77 (2015)