Two Kinds of Hypernymy Faults in WordNet: the Cases of Ring and Isolator*

Yang Liu, Jiangsheng Yu, Zhengshan Wen, and Shiwen Yu

Institute of Computatinal Linguistics, Peking University
100871 Bejing, China
Email: liuyang@pku.edu.cn, yujs@pku.edu.cn, wenzs@pku.edu.cn, yusw@pku.edu.cn

Abstract. Hypernymy is the key relation that serves to form the ontology of the noun and verb concepts in WordNet and provides a common way of making induction along the hypernymy tree for the NLP researchers. Howerver, we find 2 kinds of abnormal hypernymy in WordNet 2.0, the cases of ring and isolator for short, which can largely harass the reasoning and eventually lead to errors.

1 Introduction

As the mostly used MRD for semantic analysis nowadays, WordNet features the following items. First, the founders at Princeton University originally defined the rather abstract concept, *Concept*, by a less abstract concept, *SynSet*, which makes a *Concept* formally representable by itself. Second, they further described many kinds of relation between all these *SynSets*, which makes a *Concept* actually significative in such a semantic network.

By means of this particular organization of WordNet, the NLP researchers can, somehow, evaluate the sense of a word or phrase in its context and the *Concept* eventually emerges. The reasoning of ontology, say induction and deduction, thus gets involved.

The credibility of the reasoning lies in the description of the *Concepts* in WordNet. What really counts is that whether or not all the *SynSets* and their relations are well formed (Liu, 2002).

The relations WordNet now applied to the noun and verb *concepts* are synonymy, antonymy, hypernymy, holonymy, entailment, cause and etc., among which synonymy and hypernymy are the most important. Synonymy and hypernymy help to form the *SynSets* and their hierarchies respectively. The hypernymy tree, as the hierarchy of *Concepts*, provides a common way of making induction for the NLP researchers.

According to the specification of WordNet, the noun and verb *concepts* fall into 40 semantic categories with the noun *concepts* ranging from 04 to 28 and the verb *concepts* ranging from 29 to 43. Each category actually denotes a hypernymy tree by the hypernymy relation and its name and content list below (Fellbaum, 1999).

2 Why the Cases of Ring and Isolator Are Abnormal

In principle, hypernymy indicates the uniqueness of induction by its definition and the hypernym of a *Concept* should always be in the same category of the *Concept* proper. This is

* This research is supported by National Project 973, No.G1998030507-4 National Project 863, No. 2002AA117010-08 and Beijing Natural Science Foundation, No. 4032013.

Petr Sojka, Karel Pala, Pavel Smrž, Christiane Fellbaum, Piek Vossen (Eds.): GWC 2004, Proceedings, pp. 347–351.

Table 1. Semantic categories of the noun concepts in WordNet

Category	Name	Contents
04	Act	Nouns denoting acts or actions
05	Animal	Nouns denoting animals
06	Artifact	Nouns denoting man-made objects
07	Attribute	Nouns denoting attributes of people and objects
08	Body	Nouns denoting body parts
09	Cognition	Nouns denoting cognitive processes and contents
10	Communication	Nouns denoting communicative processes and contents
11	Event	Nouns denoting natural events
12	Feeling	Nouns denoting feelings and emotions
13	Food	Nouns denoting foods and drinks
14	Group	Nouns denoting groupings of people or objects
15	Location	Nouns denoting spatial position
16	Motive	Nouns denoting goals
17	Object	Nouns denoting natural objects "not man-made"
18	Person	Nouns denoting people
19	Phenomenon	Nouns denoting natural phenomena
20	Plant	Nouns denoting plants
21	Possession	Nouns denoting possession and transfer of possession
22	Process	Nouns denoting natural processes
23	Quantity	Nouns denoting quantities and units of measure
24	Relation	Nouns denoting relations between people or things or ideas
25	Shape	Nouns denoting two and three dimensional shapes
26	State	Nouns denoting stable states of affairs
27	Substance	Nouns denoting substances
28	Time	Nouns denoting time and temporal relations

quite true of the general linguistics theory. We, however, live in a world of reality other than theory. There do exist case that it is hard to reach the uniqueness of induction for a certain *Concept* and we can only adopt such a belief that this *Concept* might have more than one hypernym, one in its own category (the main category) and the others in other categories (the less important categories). This is an exception to the definition.

In other words, if we use H_{in} to measure the hypernyms of a certain *Concept* C_x in its own category and H_{out} to measure its hypernyms in other categories, the cases we can adopt should satisfy the condition of $0<H_{in}\leq 1$ and the value of H_{out} does not matter too much.

Then what happens to the cases not satisfying this condition? What is the meaning of these cases and whether or not this will happen in WordNet 2.0, the latest version of WordNet family by now?

The denial of $0<H_{in}\leq 1$ might be either (1) $H_{in}\geq 2$, case 1 for short, or (2) $H_{in}=0$, case 2 for short. As the root of the hypernymy tree also satisfies the condition of case 2, we strengthen the condition of case 2 by adding $H_{out}\geq 1$ to it and then get (3) $H_{in}=0$ and $H_{out}\geq 1$, case 3 for short.

(1) For case 1, $H_{in}\geq 2$ means that the current *Concept* C_x has at least 2 fathers in its own category. According to the specification of WordNet we've mentioned above, each

Table 2. Semantic categories of the verb concepts in WordNet

Category	Name	Contents
29	Body	Verbs of grooming, dressing and bodily care
30	Change	Verbs of change of size, temperature, intensity, etc.
31	Cognition	Verbs of thinking, judging, analyzing, doubting, etc.
32	Communication	Verbs of telling, asking, ordering, singing, etc.
33	Competition	Verbs of fighting, athletic activities, etc.
34	Consumption	Verbs of eating and drinking
35	Contact	Verbs of touching, hitting, tying, digging, etc.
36	Creation	Verbs of sewing, baking, painting, performing, etc.
37	Emotion	Verbs of feeling
38	Motion	Verbs of walking, flying, swimming, etc.
39	Perception	Verbs of seeing, hearing, feeling, etc.
40	Possession	Verbs of buying, selling, owning, and transfer
41	Social	Verbs of political and social activities and events
42	Stative	Verbs of being, having, spatial relations
43	Weather	Verbs of raining, snowing, thawing, thundering, etc.

category already denotes a hypernymy tree by the hypernymy relation. This condition will unavoidably lead to the case of ring in WordNet. Along these upward arcs of hypernymy of *Concept* C_x, there naturally exists C_x's most nearby ancestor, say *Concept* C_z, which has at least 2 chirldren, say *Concept* C_{y1} and C_{y2}; at the same time, *Concept* C_{y1} and $C_{y2}$are all C_x's ancestors. As WordNet is an inheritance system (Fellbaum, 1999), we can now infer that C_x shares C_{y1} and C_{y2}'s all properties, among which a pair of properties must be opposite for C_{y1} and C_{y2} have the same father C_z and hereby is distinguishable. This is paradoxical by the general linguistic theory.

(2) For case 1, H_{in}=0 means that the current *Concept* C_x has no father at all and it can be the root of the hypernymy tree. This condition doesn't lead to any faults.

(3) For case 3, H_{in}=0 and $H_{out} \geq 1$ means that the current *Concept* C_x has nothing, by the hypernymy relation, to do with any available *Concept* C_z as its father in its own category. Also, C_x has at least 1 father in other categories and actually belongs to those categories. This is nonsense and leads to the case of isolator.

In the final analysis, both the cases of ring and isolator are abnormal.

3 The Searching Algorithm and the Obtained Results

In order to explore the actual cases of ring and isolator in WordNet 2.0, we devised the searching algorithm for the noun *Concepts* demonstrated as follows. It can also apply to the verb *Concepts* after minor modification of the value of the boundary information about the semantic categories.

```
Case_Ring_Total=0
Case_Isolator_Total=0
Case_Ring_by_Category(4..28)=0
```

```
Case_Isolator_by_Category(4..28)=0
Boundary(4..28)=Begin_Offset_of_Category
Boundary(29)=Biggest_Offset_of_Dat_File
Dat_File.Recordset.MoveFirst
Do UntilDat_File.Recordset.EOF
  Number_of_IN_Hypernyms=0
  Number_of_OUT_Hypernyms=0
  Dat_File_Line_String=Data.Recordset.Fields("Dat_File_Line")
  Category=Val(Mid(Dat_File_Line_String,10,2))
  Position=InStr(Dat_File_Line_String,"@")
  Do While Position>0
    Hypernym_String=Mid(Temp,Pos+2,8)
    If Hypernym_String is between Boundary(Category)
    and Boundary(Category+1) Then
      Number_of_IN_Hypernyms=Number_of_IN_Hypernyms+1
    Else
      Number_of_OUT_Hypernyms=Number_of_OUT_Hypernyms+1
    End If
    Position=InStr(Pos+18,Dat_File_Line_String,"@")
  Loop
  If Number_of_IN_Hypernyms>=2 Then
    Record the current Dat_File_Line_String as an example of Case Ring
    Case_Ring_by_Category(Category)=Case_Ring_by_Category(Category)+1
    Case_Ring_Total=Case_Ring_Total+1
  End If
  If Number_of_IN_Hypernyms=0 and Number_of_OUT_Hypernyms>=1 Then
    Record the current Dat_File_Line_String as an example of Case Isolator
    Case_Isolator_by_Category(Category)=Case_Isolator_by_Category(Category)+1
    Case_Isolator_Total=Case_Isolator_Total+1
  End If
  Dat_File.Recordset.MoveNext
Loop
```

By this algorithm, we found 1,839 occurrences out of a total of 79,689 noun *Concepts* and 17 occurrences out of a total of 13,508 verb *Concepts* for the case of ring in WordNet 2.0. The percentages are 2.31% and 0.13% respectively. Table 3 and 4 show the detailed portion for each category.

Table 3. Cases of ring in the noun Concepts

[C04] 73	[C05] 27	[C06] 258	[C07] 12	[C08] 23
[C09] 29	[C10] 67	[C11] 5	[C12] 11	[C13] 24
[C14] 34	[C15] 205	[C16] 0	[C17] 11	[C18] 682
[C19] 7	[C20] 29	[C21] 10	[C22] 8	[C23] 13
[C24] 2	[C25] 4	[C26] 102	[C27] 193	[C28] 8

For the case of ring, there are 2,654 occurrences out of a total of 79,689 noun *Concepts* and 1,551 occurrences out of a total of 13,508 verb *Concepts* in WordNet 2.0. The percentages are 3.33% and 11.48% respectively. Table 5 and 6 show the details.

Table 4. Cases of ring in the verb Concepts

[C29] 0	[C30] 5	[C31] 0	[C32] 0	[C33] 0
[C34] 1	[C35] 4	[C36] 2	[C37] 0	[C38] 1
[C39] 0	[C40] 1	[C41] 2	[C42] 1	[C43] 0

Table 5. Cases of isolator in the noun Concepts

[C04] 65	[C05] 415	[C06] 199	[C07] 30	[C08] 93
[C09] 54	[C10] 73	[C11] 15	[C12] 42	[C13] 34
[C14] 37	[C15] 351	[C16] 6	[C17] 114	[C18] 394
[C19] 33	[C20] 286	[C21] 56	[C22] 10	[C23] 15
[C24] 72	[C25] 21	[C26] 99	[C27] 112	[C28] 28

Table 6. Cases of isolator in the noun Concepts

[C29] 104	[C30] 211	[C31] 87	[C32] 136	[C33] 69
[C34] 32	[C35] 283	[C36] 43	[C37] 36	[C38] 106
[C39] 45	[C40] 76	[C41] 197	[C42] 112	[C43] 14

4 Conclusion

To sum up, the cases of ring and isolator, as 2 kinds of hypernymy faults we've found in WordNet, can largely harass the reasoning along the hypernymy tree for the NLP researchers and eventually lead to errors. In the future, some amendments should be made to solve these issues during the evolution of WordNet.

References

1. Fellbaum, C. WordNet: an Electronic Lexical Database. Cambridge, Mass.: MIT Press (1999).
2. Liu, Y., Yu, J. S. and Yu, S. W. A Tree-Structure Solution for the Development of Chinese WordNet. GWC 2002, India (2002).
3. Liu, Y., Yu, S. W. and Yu, J. S. Building a Bilingual WordNet-Like Lexicon: the New Approach and Algorithms. COLING 2002, Taipei, China (2002).
4. Pianta, P., Pala, K. VisDic – a New Tool for WordNet Editing. GWC 2002, India (2002).
5. Vossen, P. EuroWordNet: a Multilingual Database with Lexical Semantic Networks. Dordrecht: Kluwer (1998).
6. Yu, J. S. Evolution of WordNet-Like Lexicon. GWC 2002, India (2002).