EDR's Concept Classification and Description for Interlingual Representation

Hideo Miyoshi, Takano Ogino, and Kenji Sugiyama
Japan Electronic Dictionary Research Institute, LTD.
Daini-Abe Bldg., 78-1, Kanda-Sakumagashi, Chiyoda-ku, Tokyo 101, JAPAN

October 28, 1997

Abstract

This paper describes the outline of the EDR Concept Dictionary and gives some examples of interlingual representations as the semantic representations for an input sentence.

1   Introduction

Japan Electronic Dictionary Research Institute, LTD. (EDR) has developed the EDR Electronic Dictionary [EDR1996, Miyoshi *et al.* 1996] for the purpose of advanced natural language processings. It is the product of a nine-year project (from fiscal 1986 to fiscal 1994), and consists of five types of large-scale dictionaries: the Word Dictionary, the Concept Dictionary, the Bilingual Dictionary, the Co-occurrence Dictionary, and the Technical Terminology Dictionary. The EDR Electronic Dictionary integrates the relationship between lexical entries and their concepts in the form of concepts hierarchy and semantic relations, together with database of corpus from which lexical and conceptional information were extracted.

Within the EDR Electronic Dictionary, the Concept Dictionary contains information about concepts (word senses) and their relations. The Concept Dictionary consists of three types of subdictionaries, - Head Concept Dictionary, Concept Classification Dictionary, and Concept Description Dictionary. Although Concept Classification Dictionary and Concept Description Dictionary are the static database about concepts, they have been so designed that their framework can be used for interlingual representations in machine translations.

2   EDR Concept Dictionary

EDR Concept Dictionary contains 400.000 concepts listed in Japanese and English Word Dictionaries of 200,000 words each. The role of the Concept Dictionary is to provide the data required for computer processing of the semantic contents or the concepts, expressed in natural language sentences, such as:

- Generating appropriate semantic representations for sentences

- Determining the similarity (equivalence) of semantic contents

- Converting a semantic content into a similar (equivalent) content

For this reason, the Concept Dictionary contains three types of subdictionaries:

- Headconcept Dictionary

- Concept Classification Dictionary

- Concept Description Dictionary

The Headconcept Dictionary contains the headconcept and concept explication. The headconcept is a word whose meaning is close to the content of the concept. Concept explication is described in natural language to explicate the contents of the concept. The Concept Classification Dictionary contains the classification of concepts that have a super-sub relation. The Concept Description Dictionary contains the set of pairs of concepts that have certain semantic relations other than super-sub relations.

3   Concept Classification Dictionary

The Concept Classification Dictionary contains the classification of concepts that have a super-sub relation, i.e. hierarchical structure among concepts. According to the classification of concepts used in a thesaurus and other reference sources, concepts are classified according to a tree structure. (In the EDR Concept Classification Dictionary, multiple inheritance is allowed thus making it possible for one concept to have more than two super-concepts. In such a case, the classification is not actually a 'tree structure'. However, the term 'tree structure' is also used for the sake of convenience.)   The leaf
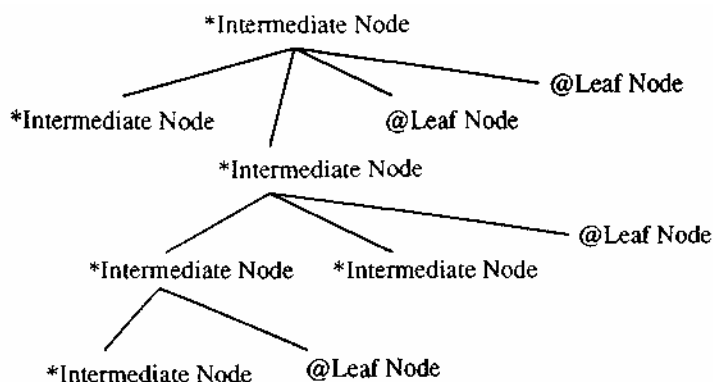
Figure 1: Intermediate and Leaf Node in the Concept Classification

nodes are always linked to the entry words of the EDR Word Dictionary, namely, they represent the word senses. Figure 1 shows the overview of the tree structure of the Concept Classification Dictionary. In Figure 1, a node marked with '*' stands for an intermediate node and a node marked with '@' stands for a leaf node. Currently the number of concept groups (intermediate nodes) is about 6,000 and the maximum depth of the tree is sixteen. In the Concept Classification Dictionary, the pair of concepts that have an immediate super and sub-concept relation are registered in one record. The concept classification is expressed by listing all the super-sub concept pairs. Figure 2 shows the upper part of the Concept Classification Dictionary.

4   Concept Description Dictionary

The Concept Description Dictionary is a set of concept description entries which are composed of two concepts linked by a certain concept (semantic) relation label. Currently, the following eight semantic relations are used:

```
object      agent    goal    implement
a-object  place     scene   cause
```

The a-object is an object of a particular attribute. The following are the examples of the concept description entries:
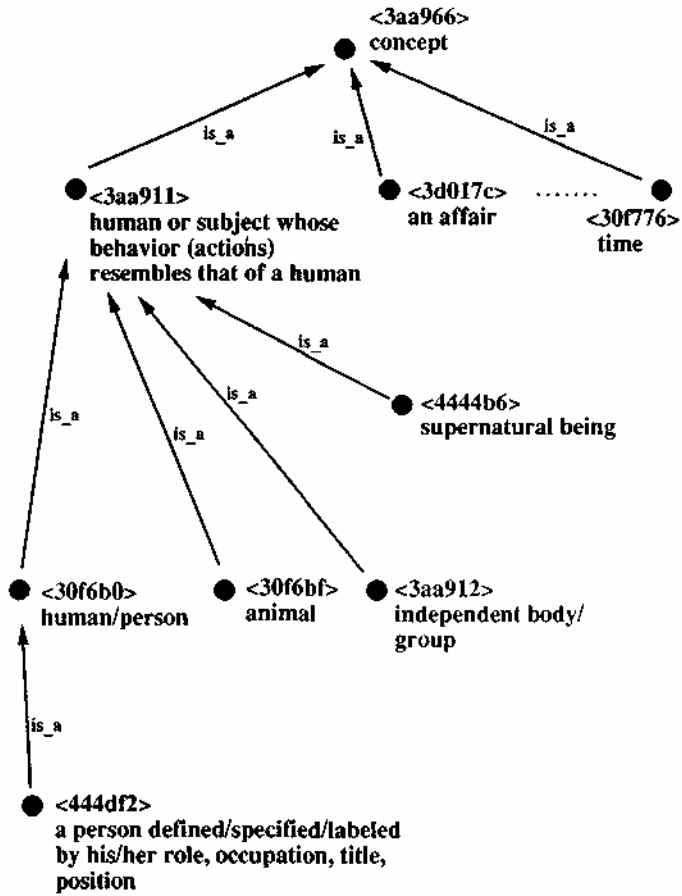
Figure 2: Upper Part of the Concept Classification Dictionary

```
<<fly>>    --   agent  -->  <<swallow>>
<<read>>   --   agent  -->  <<people>>
```

There are two types of concept description entries. The two types differ
in where the description comes from. The first type of the concept descrip-
tion type is based on human intuition, while the second type is based on the
results of analyses of actual sentences (i.e. EDR Corpus).

The Concept Description Dictionary can be used to reduce the amount
of concept description entries. For example, it is very troublesome and un-
reasonable to list up all the flying birds:

```
<<fly>> -- agent -> <<swallow>>
<<fly>> -- agent -> <<sparrow>>
<<fly>> -- agent -> <<swan>>
etc.
```

Instead of listing up all the instances, those concept description entries
can be reduced into one entry, using the upper concept "bird" of "swallow",
"sparrow", "swan", and so on:

```
<<fly>>   --   agent  ->  <<bird>>
```

5   Concept Dictionary Framework For Interlin-gual Representation

Although the EDR Concept Classification Dictionary and the EDR Concept
Description Dictionary are the sets of binary relations between two concepts,
their framework can be used for the interlingual representation, as a seman-
tic structure of a sentence. For example, the semantic representation of a
sentence 'People read books' can be expressed in two concept relations:

```
<<read>>   --   agent  -->  <<people>>
<<read>>   --object  -->  <<book>>
```

Label	Explication
agent	That which acts on its own volition and is the subject that brings about
object	That which is affected by an action or change
a-object	That which has a particular attribute
implement	That which is used in a voluntary action such as tools or other implements
material	That which is used to make up something
source	Location from which an event or occurrence begins
goal	Location from which an event or occurrence ends
place	Place (physical location) at which something occurs
scene	Place (abstract location) at which something occurs
basis	That which is used as the standard of comparison
manner	Way in which an action or change occurrs
time	Time at which something occurrs
time-from	Time at which something begins
time-to	Time at which something ends
quantity	Amount (quantity) of a thing, action, or change
modifier	Modification

Table 1: Semantic Relation Labels used in Concept Relation Representation

Between the concepts "read" and "people" there is an agent relation and between the concepts "read" and "book" there is an object relation.

Here we show some examples the semantic representations used in EDR Corpus. EDR Corpus is a tagged corpus composed of surface sentences, their constituent information, morphological information, syntactic information, and semantic information. The semantic information is called 'concept relation representation' and represents the meanings of the sentence within Concept Dictionary's framework. In addition to the eight semantic relations mentioned above, the 'concept relation representation' incorporates lots of semantic relations to express the minute semantics (Table 1).

Table 2 shows the semantic labels used to indicate aspect of the sentences.

Label	Explication
begin	Indicates an event or action starts
end	Indicates an event or action has ended
progress	Indicates an event or action is not yet completed
continue	Indicates a repetitious action/movement continues
state	Indicates the result (resulting state) of an action is continuous
yet	Indicates a pending state (the event has neither started nor ended)
already	Indicates a state/event has either started or ended
soon	Indicates a state/event will either start of end soon
just	Indicates a state/event has just started or ended
complete	Indicates the completion of an event
come	Indicates an event approaches the speaker's point of reference
go	Indicates an event gets further away from the speaker's poinnt of reference

Table 2: Labels Used to Indicate Aspect

Other semantic labels are ones used to indicate speaker's intention, ones to indicate intention, feeling or judgment of an utterance, ones for specifying/restricting concepts, ones used for representing comparison, ones used for representing numerical values, and so on.

The following are some examples of the 'concept relation representations' for actual sentences.

Example 1:

```
Sentence:
     That  airport  is  always busy.


Concept Relation Representation:
     [  [main 9:"bus":"=W crowded"]
        [attribute present]
        [object  [   [main 3:"airport":3bd5a2]
                  [attribute  that  topic]]]
        [manner 7:"always":3ceafl]]
```

Example 2:

```
Sentence:
     Test runs will begin in February.


Concept Relation Representation:
     [ [main 8:"begin":3cf349]
        [object  [ [main 3:"run":262a03]
        [attribute topic]
        [number @1:"c#2-or-more":]
        [modifier 1:"test":3ce65d]]]
        [time 12:"February":102054]]
```

Example 3:

```
Sentence:
     Testing is to continue until July 30.


Concept Relation Representation:
     [ [main 7:"continue":"=W to go on over a long period without
                              stopping"]
```

```
[which  [ [main @2:"c#schedule":]
           [object 7:"continue":"=W to go on over a long
                                   period without stopping"]]]
[attribute future continue]
[object 1:"testing":3ce65d]
[time-to  [ [main @l:"c#day":]
            [modifier 14:"30":00011e]
            [modifier 12:"July":3bce11]]]]
```

6 Conclusion

This paper presented the outline of the EDR Concept Dictionary and the "concept relation representation" within its framework as an interlingual representation. We expect this framework will provide the basic components for constructing the interlingual representations of the input sentences in interligual MT systems. EDR has been making an improvement and a refinement of the EDR Concept Dictionary in order to build it as a large-scale standard ontology[Ogino *et al*.,1997]. We hope the EDR Concept Dictionary will contribute toward a development of intelligent MT systems.

References

[EDR1996] EDR. *EDR Electronic Dictionary Version 1.5 Technical Guide.* EDR TR2-007, 1996.

[Miyoshi *et al*.1996] Hideo Miyoshi, Kenji Sugiyama, Masahiro Kobayashi, and Takano Ogino. *An Overview of the EDR Electronic Dictionary and the Current Status of Its Utilization.* Proceedings of COLING-96, August 1996.

[Ogino *et al*. 1997] Takano Ogino, Hideo Miyoshi, Fumihito Nishino, Masahiro Kobayashi, Jun'ichi Tsujii. *An Experiment on Matching EDR Concept Classification Dictionary with WordNet.* Workshop on Ontologies and Multilingual NLP, IJCAI-97, 1997.