# Constructing an Ontology of Coherence Relations: An example of 'causal relation'*

Hae-Yun Lee and Sueun Jun

Dept. of Linguistics and Cognitive Science, Hankuk Univ. of Foreign Studies, Korea

haeyun@hufs.ac.kr, kazoopablo@hotmail.com

**Abstract.** The goal of this paper is to present the methodology of constructing a language ontology of 'coherent relations'. We construct an ontology of more abstract concepts by combining an upper-level ontology and a middle-level ontology. At the former we use theoretical considerations following Sanders et al. (1992, 1993), and at the latter we use lexical items through the substitutability test of Knott and Dale (1994).

**Keywords:** ontology, coherence relations, connectives

## 1. Introduction

### 1.1.Ontology

'Ontology' has recently become one of the most attractive research areas. Even though the notion originates from philosophy, there have been many researches within AI or Computer Science. Ontology is usually defined as "an explicit specification of a conceptualization," where a "conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose" (Gruber 1993: 199). But many constructed ontologies are not consistent, because the conceptualization is different according to the interests of researchers. Therefore most ontologists make use of languages which are considered to reflect the objects of conceptualization objectively.

If we consider that an element of conceptualization is a concept, it is followed that the notion of ontology is related closely with languages. As we know, the modern linguistics is based on the 'meaning triangle' (Ogden and Richards 1923). At the triangle, the connection of a symbol with an object is mediated by a concept, and each element constructs its own system. So the system of symbols, i.e. a network of words, is similar to the system of concepts, i.e. an ontology. Therefore we can assume, that an ontology can be constructed indirectly by means of words.

The network of words is often called 'Language Ontology'. Nickles et al. (2007) considers Language Ontology as an answer to the following question: "What kinds of things do people talk as if there are?" Further the notion of Language Ontology is defined as "a conceptualization or categorization of what normal everyday human language can talk about" (Zaefferer 2002).



*22nd Pacific Asia Conference on Language, Information and Computation, pages 245–252*

### 1.2. Goals

The goal of this paper is to present the methodology of constructing the language ontology about 'coherent relations'. Some previous language ontologies such as WordNet, etc. have been restricted to the lexical categories such as nouns, verbs, adjective, and adverbs. There were no ontologies about abstract notions such as coherent relations which are realized mostly as minor categories such as suffixes or connectives.

For the construction of ontology of coherent relations, we will adopt the following strategy: At the upper-level, we will construct the ontology in a top-down manner, reflecting the theoretical considerations which are relatively language-independent. But at the middle-level, we will proceed with the work in a bottom-up manner. That is we will construct the middle part of the ontology by investigating lexical items. Since the construction of all the coherent relations is an enormous project, we will focus on the 'causal relations' in this paper.

With the ontology to be constructed we can make a typological analysis about coherence relations. For example we can compare the Korean ontology with the English or the Dutch ontology presented in Knott (1996). The trend of typological researches is based on the so-called 'ontolinguistical' approach.[1]

## 2. Coherence

### 2.1. Coherence relations

Discourse is more than a random set of utterances, and shows connectedness. The connectedness is captured by the concept of 'cohesion' and 'coherence'. In comparison with cohesion, coherence is an abstract concept which language users establish by relating the different information units in the text.

Coherence is divided into 'referential coherence' and ' relational coherence' (Sanders and Maat 2006). The latter has been investigated under the theme 'coherence relations'.[2] That is, under the assumption, that the interpretation of the related segments needs to provide more information than is provided by the sum of the segments taken in isolation, text grammarians adopt the view that text segments are connected by coherence relations like CAUSE-CONSEQUENCE between them. Such coherence relations can be made explicit by linguistic markers, so-called connectives or cue phrases, but not always, as we see in (1).

(1) (a) The buzzard was looking for prey. The bird was soaring in the air for hours.
(b) Gareth grew up during the 1970s, *so* he loves disco music.

It has been disputed, what coherence relations are. Some have insisted that coherence relations should be considered as cognitive entities (Hobbs 1979, Mann and Thompson 1988, Sanders et al. 1992, 1993). In addition to that, we can find a similar view in Lyons (1977). He divided entities into 3 subtypes: first-order entities, second-order entities, and third-order entities. He took the third-order entities to be "such abstract entities as proposition, which are outside space and time" (ibid. 443p.). Furthermore, he mentioned the possibility of distinction in those entities, for example between psychological and non-psychological entities. We think that coherence relations in the sense of Sanders et al. (1993) can correspond to those psychological third-order entities.

[1] "the most reliable basis for any cross-linguistic research lies in the common core of the different individual human ontologies. This is the basic tenet of all approaches that can properly be called ontology-based linguistics or *ontolinguistics* for short." (Schalley and Zaefferer 2007: 3)

[2] The notion is also called as 'rhetorical relations' (Mann & Thompson 1988), 'clause relations', 'discourse relations', …

There is no agreement among researchers about the kinds/type of relations and the number of relations. For example, Hobbs (1978), Mann and Thompson (1988) etc. hypothesize dozens of relations (cf. Hovy 1990).

With respect to 'causal relation', Mann and Thompson (1988) presented a more detailed classification by considering speaker's volition and ordering between two clauses: *volitional cause, non-volitional cause, volitional result, non-volitional result, purpose*.

## 2.2. Linguistic realizations

We can sometimes identify the coherence relations at the surface form. In such cases, coherence relations can be realized at the sentence level and at the discourse level.

As for the sentence level, some adverbial/subordinate clauses show the realization of coherence relations. According to Thompson and Longacre (1985), languages of the world use 3 devices to mark subordinate clauses: (i) subordinating morphemes, (ii) special verb forms, (iii) word order. As for the discourse level, discourse markers or cue phrases are typical means to represent coherence relations, by which sentences are connected to each other. For example, phrases such as *as a result, therefore* etc. belong to this category.

In Korean, there are two groups of lexical items which realize coherence relations. The first group is a list of suffixes which are attached to the verbal stems of subordinate clauses. The concrete coherence relation is determinded by the suffix to be attached. The following shows the difference of verbal suffixes at the main clause and at the subordinate clause.

(2) (a) ku-nun yelsimhi kongpuha-yss-*ta*.
He-TOP hard study-PAST-DEC
'He studied hard.'

(b) ku-nun yelsimhi kongpuha-yss-*ciman* sihem-ey tteleci-ess-ta
He-TOP hard study-PAST-although test-ACC fail-PAST-DEC
'Although he studied hard, he failed the test.'

The second group is a list of connectives which relate two clauses at the discourse level. The example in (3) shows that type.

(3) ku-nun yelsimhi kongpuha-yss-ta. *kulemeyto* sihem-ey tteleci-ess-ta.
He-TOP hard study-PAST however test-ACC fail-PAST-DEC
'He studied hard. However, he failed the test.'

# 3. Language Ontology

## 3.1. Upper-level Ontology

As we mentioned above, there are a few ontologies about coherence relations. But strictly speaking, the ontologies are merely taxonomies which include fewer levels and the smaller number of nodes in the hierarchy. Therefore we will try to construct an ontology in the original sense.

As for the upper-lever of the ontology, there is an interesting research which has considered some philosophical discussions. Sanders et al. (1992, 1993) presented the classification of coherence relations, based on more elementary notions. Following Sanders et al. (1992, 1993), we will construct an upper-level ontology. At first we look at their classification. They hypothesized four basic notions, each of which can take two alternative values.

- Basic Operation: CAUSAL/ADDITIVE
  CAUSAL relations are those where a 'relevant' causal connection exists between the spans; all other relations are ADDITIVE.

- Source of Coherence: SEMANTIC/PRAGMATIC
  It is SEMANTIC if the spans are related in terms of their propositional content and PRAGMATIC if they are related because of their illocutionary force.
- Polarity: POSITIVE/NEGATIVE
  A relation is POSITIVE if its basic operation links the content of the two spans as they stand, and NEGATIVE if it links the content of one of the spans to the negation of the content of the other span.
- Order of Segments: BASIC/NON-BASIC
  CAUSAL relations are deemed to have BASIC order if the antecedent is on the left, and NON-BASIC order if it is on the right.

Of 4 basic notions, the notion 'Order of Segment' is related with the superficial realization. So if we neglect the notion and revise the taxonomy of Sanders et al. (1993), we can present the upper-level ontology as follows.

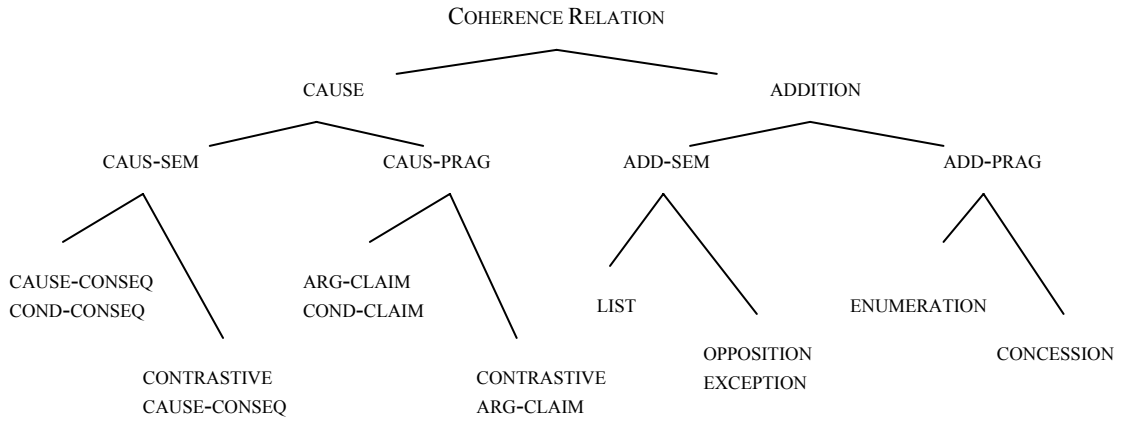


**Figure 1**: An upper-level ontology

At first, we can divide coherence relations into 'cause' and 'addition' according to the 'source' notion. Next the 'cause' is divided into a semantic concept 'cause-sem' and a pragmatic concept 'caus-prag' according to the 'source' notion. Finally using the 'polarity' notion, we can divide again the former into 'cause-consequence/ condition-consequence' (4ab) and 'contrastive cause-consequence' (4c), and the latter into 'argument-claim/ condition-claim' and 'contrastive argument-claim'. We can see the corresponding examples below (Sanders et al. 1993):

(4) (a) Because there is a low-pressure area over Ireland, the bad weather is coming our way.
(b) Ready? Then we're now off on safari.
(c) Although the number of similarities between faces is enormous, we do not have the slightest difficulty in distinguishing a very large number of people.

## 3.2. Middle-level Ontology

As we have seen above, the upper-level ontology is constructed by philosophical or psychological considerations. So the level of the hierarchy and the number of concepts are sparse. In this section we try to construct a richer middle-level ontology which is connected with the upper-level ontology in Fig. 1. To simplify the explanation, we restrict ourselves to the concept 'cause-sem'.

As is well known, the most explicit markers signaling coherence relations are connectives.[3] Although there is no one-to-one correspondence between coherence relations and connectives, we can construct the middle-level ontology by examining the distribution of connectives. Knott (1996) and Knott and Dale (1994) presented taxonomy of connectives in English and in Dutch. As a starting point, let us review their methodology.

### 3.2.1. Knott and Dale (1994)

Knott and Dale (1994) classified cue phrases according to their syntactic properties. This classification is made by a simple linguistic test, so-called 'substitutability'. The test calls for the judgment of a writer, if a cue phrase can be replaced by another cue phrase.

(5) The bouncers refused us access to the bar, { *because* / ✓ *on the grounds that* / # *therefore* } we were wearing jeans.

In (5), *on the grounds that* is represented as substitutable for the original cue phrase *because*, whereas *therefore* is not substitutable for the cue phrase. Seeing concretely, there are 4 substitutability relationships (Knott and Sanders 1998):

- X is synonymous with Y if in any context where one can be used, the other can also be used.
- X and Y are exclusive if they can never be substituted for one another in any context.
- X is a hypernym of Y if whenever Y can be used, so can X; but there are some contexts where X can be used and Y cannot.
- X and Y are contingently substitutable if there are some contexts where they can be substituted, other contexts where X can be used and not Y, and still other contexts where Y can be used and not X.

Based on that 'substitutability test', they presented the taxonomy of cue phrases in English and Dutch. But we cannot construct ontology by means of all the 4 relationships. In constructing an ontology, we have no means to represent 'exclusion' and 'a contingent substitutability' differently. Therefore we will merge the two relationships into one, and will use in total 3 relationships in the construction of Korean ontology.

### 3.2.2. A Korean Ontology

Next, we illustrate the construction of middle level ontology, using the substitutability test of Knott and Dale (1994).

As we have seen above, coherence relations in Korean are realized by suffixes and connectives. For the test we collected such cue phrases from different sources. As for connectives, we selected lexical items signaling causal relations from Im et al. (2001), where the whole list of Korean connectives is presented. The size of selected connectives is 28 tokens which correspond to 20 types. In the case of suffixes, there are no researches which present the whole list. Therefore we collected the items from the tagged corpus 'Sejong corpus'. In details we extracted all the items which are tagged with 'connecting'-suffixes, and selected the suffixes showing the causal relation from the items. By this procedure we have gotten 31 connectives (19 types). The table 1 shows the whole list we got.

**Table 1:** The list of Korean cue phrases

[3] They are also called as 'discourse/ coherence/ lexical markers', 'discourse operators', 'discourse/ pragmatic/ sentence connectives', 'cue phrases', 'clue words', etc.

| Connectives | Suffixes |
|---|---|
| kulayse, kulayya, kulehani/kuleni, kulehancuk/kulencuk/kulihancuk, kulenmankem/kulenimankum/kulenimanchi, kulemulo, kuleca, kulenkolo, kulihaye, ilihaye, iey, kyelkuk, ttalase, hanun-su-epsi, ilehkeytoyca, kulena, kulentey/kulenteto, kulemeyto, kulehciman/kulehcimanun, haciman, kulemeyto pulkuhako | -ase/-ese, -myen/-myenun/-myenya, -ni/-nika, -ulsulok, -tamyen, -mulo, -nula/-nulako, -killay, -cani, -teni/-teniman, -koseya, -nunpa, -may, -layse, -uncuk, -nolani/-nolamyen/-nolanika, -koto, -nunteyto, -ciman/-cimanun |

Now we turn to the construction of ontology through the substitutability test. The middle-level ontology to be constructed at this section will be connected to the concept 'CAUSE-CONSEQUENCE/ CONDITION-CONSEQUENCE' at the upper-level ontology (Fig. 1). This methodology can be applied to other concepts of Fig. 1 in a similar way.

As a starting point, we follow Chang (1995) where the suffix '-myen' is considered as the most general suffix among the suffixes representing a causal relation. In addition to '-myen', there are suffixes '-nula'/'-killay' which represent a causal relation, but behave themselves differently. As we see below, the latter group expresses a causal relation on the basis of a temporal connection between events.[4]

(6) (a) syawe-lul ha-*myen* kipun-i sangkway-ha-ta.
shower-ACC do-SUFF feeling-NOM refresh-do-DEC
'If you take a shower, then you feel refreshed.'
(b) # syawe-lul ha-*nula* kipun-i sangkway-ha-ta.
shower-ACC do-SUFF feeling-NOM refresh-do-DEC
'I feel refreshed to take a shower.'
(c) syawe-lul ha-*nula* cenhwa-lul mos-pat-ass-ta
shower-ACC do-SUFF phone-ACC not-take-PAST-DEC
'I couldn't answer the phone, because I took a shower.'

Let us examine the suffix '-myen' in details. If we apply the test of Knott (1996) to Korean, it is revealed that there are subordinate connectives '-kulayse' and '-lyeko' under the connective '-myen'. The tests below show such a sub-classification.

(7) (a) 1-e 2-lul teha-*myen* 3-i-ta.
1-ACC 2-OBJ plus-SUFF 3-COP-DEC
'1 plus 2 is 3.'
(b) 1-e 2-lul teha-yss-ta. *kulayse* 3-i-ta.
1-ACC 2-OBJ plus-PAST-DEC. so 3-NOM-DEC
'I plus 1 to 2. so it is 3 now.'
(8) (a) kongpu-lul yelsimhi ha-*myen* cohun sengcek-ul et-unu-ta.
study-OBJ hard do-SUFF good grade-OBJ get-PRES-DEC
'If you study hard, you will get good grade.'
(b) cohun sengcek-ul et-*ulyeko* kongpu-lul yelsimhi ha-yss-ta.
good grade-OBJ get-SUFF study-OBJ hard do-PAST-DEC
'To get the good grade, I studied hard.'

[4] According to Yim (1999), suffixes '-nula' and '-killay' show different syntactic distributions. But such differences are assumed to be recorded under the individual item.

From this observation we can conclude that there are sub-concepts such as ‘cause’ with ‘-kulayse’ and ‘purpose’ with ‘-lyeko’ under the concept ‘condition’ with ‘-myen’. Furthermore, the following tests show that the two sub-concepts are inter-exclusive each other.

(9) (a) √ pay-ka kop-ass-ta. *kulayse* pap-ul mek-ess-ta.
stomach-NOM hungry-PAST-DEC so meal-OBJ eat-PAST-DEC
‘I was hungry. So, I had a meal,’
(b) # pap-ul mek-*uleko* pay-ka kop-ass-ta.
meal-OBJ eat-SUFF stomach-NOM hungry-PAST-DEC
‘I was hungry to had a meal..’

(10) (a) √ pap-ul mek-*uleko* sang-ul chali-ess-ta.
meal-OBJ eat-SUFF table-OBJ set-PAST-DEC
‘I set the table to eat.’
(b) #/? sang-ul chali-ess-ta. *kulayse* pap-ul mek-ess-ta.
table-OBJ set-PAST-DEC. so meal-OBJ eat-PAST-DEC
‘I set the table. That’ why I had a meal.’

Finally the concept ‘cause’ can be specified into a more concrete concept, as we see in (11). That concept is realized by the connectives such as ‘kyelkwuk’, ‘hanun swu epsi’, etc.

(11) (a) pay-ka kop-ass-ta. *kulayse* pap-ul mek-ess-ta.
stomach-NOM hungry-PAST-DEC so meal-OBJ eat-PAST-DEC
‘I was hungry. So, I had a meal,’
(b) pay-ka kop-ass-ta. *kyelkuk* pap-ul mek-ess-ta.
stomach-NOM hungry-PAST-DEC. finally meal-OBJ eat-PAST-DEC
‘I was hungry. Finally, I had a meal,’

From the tests above, we can get the following middle-level ontology which is connected with the concept ‘cause-consequence’/’condition-consequence’ of the upper-level ontology (Fig. 1).

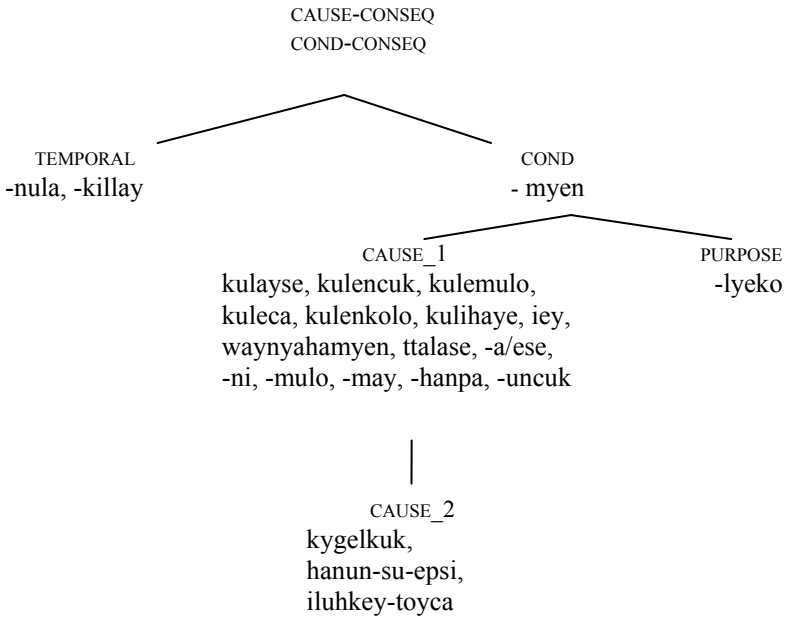


**Figure 2**: The middle-level ontology

## 4. Conclusion

In this paper, we presented the method of constructing an ontology of coherence relations. We could construct an ontology of more abstract concepts by combining an upper-level ontology and a middle-level ontology. In case of the former we used the theoretical considerations following Sanders et al. (1992, 1993), and in case of the latter we used lexical items through the substitutability test of Knott and Dale (1994).

The resulted Korean Ontology can be used for the typological analysis of coherence relations, because the comparison of languages tends to be based on the ontology. In future we will extend the ontology into the whole coherence relations and try to do a typological analysis between different languages.

## References


Chang, Kyeon-hee. 1995. The semantic structure of Korean connective endings. *Hangul 227, 151-174*

Gruber, Thomas R. 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition Journal* 5:199-220.

Hobbs, Jerry R. 1978. Why is Discourse Coherent? In *Technical Note*. Menlo Park CA.

Hobbs, Jerry R. 1979. Coherence and coreference. *Cognitive Science* 3:67-90.

Hovy, Eduard H. 1990. Parsimonious and profligate approaches to the question of discourse structure relations. Paper presented at *the 5th International Workshop on Natural Language Generation*, Pittsburgh.

Im, Yoo-jong. Pak, Dong Ho. and Hong, Chai-Song. 2001. A study on connective adverbs in Korean. Linguistics 28, 177-208.

Knott, A. 1996. A Data-Driven Methodology for Motivating a Set of Coherence Relations, Dept. of Artificial Intelligence, Uni. of Edinburgh.

Knott, A., and Sanders, T. 1998. The Classification of Coherence Relations and their Linguistic Markers: An Exploration of Two Languages. *Journal of Pragmatics* 30:135-175.

Knott, Alistair, and Dale, Robert. 1994. Using Linguistic Phenomena to Motivate a Set of Coherence Relations. *Discourse Processes* 18:35-62.

Lyons, John. 1977. *Semantics*.vol. 2. Cambridge: Cambridge University Press.

Mann, William C., and Thompson, Sandra A. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8:243-281.

Nickles, Matthias, Pease, Adam, Schalley, Andrea C., and Zaefferer, Dietmar. 2007. Ontologies across disciplines. In *Ontolinguistics -How Ontological Status Shapes the Linguistic Coding of Concepts*, eds. Andrea C. Schalley and Dietmar Zaefferer, 23-67. Berlin, New York: Mouton de Gruyter.

Ogden, C.K., and Richards, I.A. 1923. *The Meaning of Meaning*. [n.p.]: [n.p.].

Sanders, T., Spooren, W., and Noordman, L. 1992. Toward a taxonomy of cohrence relations. *Discourse Processes* 15:1-35.

Sanders, T., Spooren, W., and Noordman, L. 1993. Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics* 4:93-133.

Sanders, T., and Maat, H. Pander. 2006. Cohesion and Coherence: Linguistic Approaches. In *Encyclopedia of Language & Linguistics*, ed. Keith Brown. Amsterdam et al.: Elsevier.

Schalley, Andrea C., and Zaefferer, Dietmar. 2007. Ontolinguistics - An outline. In *Ontolinguistics - How Ontological Status Shapes the Linguistic Coding of Concepts*, eds. Andrea C. Schalley and Dietmar Zaefferer, 3-21. Berlin, New York: Mouton de Gruyter.

Thompson, Sandra A., and Longacre, Robert E. 1985. Adverbial clauses. In *Language typology and syntactic description: complex constructions*, ed. Timonthy Shopen, 171-234. Cambridge: Cambridge Univ. Press.

Yim, Eun Ha. 1999. A Semantic Analysis of Causal Conjunctive Endings in Korean. *The International Association for Korean Language Education 9th conference, 191-202.*