Semantic Representation and Composition for Spatial Concepts in Extended-HowNet

Yueh-Yin SHIH, Shu-Ling HUANG, Yi-Jun CHEN , Keh-Jiann CHEN

Institute of Information Science
Academia Sinica, Taipei

***Abstract*-We distinguish two different types of spatial concepts, places and locations in Extended-HowNet by illustrating the representation of place nouns, place adverbs, and place prepositions. For a lexical knowledge representation system, it is necessary to encode both relational senses and content senses for each word. We add new features to the HowNet definitions for function words to represent concepts more precisely such that semantic composition process can be carried out while composing words into a phrase and phrases into a sentence. The new features include fine-grain semantic roles and their taxonomy. New event roles and directional functions are employed to define spatial concepts.**

I. INTRODUCTION

Using lexical semantic knowledge to solve natural language processing (NLP) problem is getting popular in recent years. To build a semantic lexical database which provides word sense information, such as different senses of a word, synonymous or hyperonymy relations between words, etc. becomes an urgent task at the same time. HowNet is an on-line common-sense knowledge base that provides the inter-conceptual relations and attribute relations of concepts for lexical senses of the Chinese language and their English equivalents [1]. Comparing with WordNet [2], which is the most famous semantic lexical database currently, the advantage of HowNet is that each concept class can link to other concepts in various ways, not limited to explicit links, such as hyponymy relations. For instance, we can easily find the kill-agent group in HowNet by searching the definition of {X:{kill: agent={~}}}. We can find the examples of killer, strangler, assassin, butcher, and King of Hell etc.. The flexibility of defining word senses makes HowNet a better ontological representation tool in NLP field [3].

From 2003, we adopt the HowNet-based meaning representation mechanism to define word meanings of over 70,000 lexical entries in the CKIP (Chinese Knowledge Information Processing) Chinese Lexical Knowledge Base. In addition, we invented and added the description of Multi-level meaning representation and complex relation to form a new knowledge representation model called Extended-HowNet. Each concept is defined by simpler concepts instead of semantic primitives only, and complex definitions are facilitated to define kinship relations, temporal and spatial concepts and comparative notions etc. [4]. Extended-HowNet aims to differentiate the ambiguous senses and denote the functions of lexicons (such as the different functions between function words and content words). Besides, Extended-HowNet also contributes to meaning composition in phrases and sentences automatically in computer systems. In this paper, we'll take spatial concepts, including place adverbs, place nouns, and place prepositions as examples to illustrate advantages of our designed Extended-HowNet. Those lexicons are all spatial-sense related concepts with differences in content senses and relational senses. The representations to differentiate senses show the advantages of the Extended-HowNet.

In section 2, some related work is discussed. Section 3 presents the way of defining different spatial concepts in Extended-HowNet. Examples of concept composition use definitions of spatial concepts are shown in Section 4. Conclusion and future work are given in Section 5.

II. BACKGROUNDS

The concept representational mechanism of Extended-HowNet, which is evolved from HowNet, is stated as follows. In general, in HowNet a lexical sense is defined by its hypernym concept and differentiation features. For example, the concept 'white clouds' is defined as (1):

(1) *baiyun*
白雲
white clouds
def:{CloudMist|雲霧:color={white|白}}

The semantic role 'color' describes a binary relation between concepts of 'CloudMist' and 'white' to differentiate 'white clouds' from other kind of clouds. However, some types of concepts do not have natural hypernym concepts. For example, the parts of an object usually have no natural hypernym concepts. Instead, they are linked to other concepts by part-whole relations in the ontology. We extend Extended-HowNet with a new notation, %, to denote *part-of*, and use the feature relations of the location or telic to distinguish different parts. For example:

(2) *jiao*
腳
foot
def:{%animal|獸:telic={walk|走:agent = {~}}

Besides, the definitions of relational-type concepts, such as kinship relations and directional relations, are different from

0-7803-9361-9/05/$20.00 ©2005 IEEE

the definitions of entities. For instance, *zufu* 'grandfather' and *xibeijiao* 'north-west suburb' have to be expressed by composing unary relations, instead of feature attributes [1].

(3) *zufu*
祖父
grandfather
def: {father(father(human:x))}

(4) *xibeijiao*
西北郊
north west suburb
def: {place|地方=north(west(suburb|郊區))}

In (4), both 'north' and 'west' are regarded as place to place functions for denoting the detail direction of an entity. Many spatial concepts are defined by those place to place function features, the definitions can be formulized as a=function(x).

Function words, which add mainly grammatical information, establish the relationship between two concepts. In contrast to content words, function words contain less content senses, but rich relational senses. In HowNet, function words are all defined as {FuncWord|功能詞: semantic role={}}. Definitions of this kind can just refer to their parts-of-speech but fail to provide semantic information. Although there is no sharp distinction between function words and content words, function words do play mainly grammatical functions, we therefore modify the Extended-HowNet definition mechanism of function words to encode both relational senses and content sense of function words. As a result of this change, the unification process during the semantic composition can achieve under the Extended HowNet representation model [5]. Next section focuses on the definition mechanism of three different types of spatial concepts.

III. Representation Of Spatial Concept At Extended-Hownet

Two different types of spatial concepts are distinguishable here. The first type is concepts of places, such as *Xi ban ya* 'Spain', *shichang* 'market', *zuobian* 'left-hand side', and *dong nanfang* 'the South East'. The second type is concept of locations, such as *yanjie* 'along the street', *daochu* 'everywhere', and *dang chang* 'on the spot', which denote places where events occurred, i.e. they also link events to occurring places. In the following sections, differences of these two types of concepts will be illustrated by their Extended-HowNet representations.

A. Place Nouns

Place nouns are substantives that express spatial concepts [6]. If a place noun is the name of a specific place, it is a proper noun and represented as (5) in Extended-HowNet.

(5) Taipei
def:(city|城市:name= 'Taipei',
location=(Taiwan|台灣))

We use a round bracket to express an individual in order to differentiate from a generic concept, which uses a curly bracket and signifies a nonspecific member of a group. For example, we describe 'study room' and 'river' as (6), (7)

(6) *shufang*
書房
study room
def:{room|房間:telic={study|學習:
place|地方={~}}}

(7) *he*
河
river
def: {waters|水域:modifier={linear|線}}

Some place nouns are nominal-localizer compounds, which indicate a specific part of place. To denote a specific part of place is similar to describe a part-whole relation. There is no hypernym concept can be used as major sense. Therefore directional functions are adopted to describe some specific parts of places and functional compositions are also used to describe complex directions. When we need to describe the detail of a place, we then have to use these complex relation definition forms. Therefore we define 'inside the study room' and 'river bank' as (8), (9)

(8) *shufang nei*
書房 內
inside the study room
def: {place|地方=internal(study room|書房)}

(9) *hebian*
河邊
river bank
def: {place|地方=side(river|河)}

Both internal(study room) and side(river) are examples for unary relation expression. Other complex definitions examples are:

(10) *neiye*
內野
the infield
def:{place|地方=
internal(baseball ground|棒球場)}

(11) *shushang*
樹上
treetop
def: {place|地方=upper(tree|樹)}

(12) *dongnanfang*
東南方
south east
def: {place|地方= south(east())}

B. Place Adverbs

Place adverbs are event modifiers that indicate place where event occurred. They can answer the where questions. Like other types of adverbs, place adverbs are chiefly used to modify a verb, adjective, or another adverb syntactically and play thematic roles of location that indicate the place where an entity is connected with an event. For example, in the sentence (13):

(13) *ta* *yilushang* *dou* *ku* *zhe*
她 一路上 都 哭 著

She was crying throughout the journey.

The place adverb 'throughout the journey' modified the verb 'cry'. At the same time, it indicated the place where the subject 'she' connected with the event 'cry'. The latter obviously provides us more semantic information about place adverbs. Therefore, we'll define the place adverb 'throughout the journey' as (14):

(14) *yilushang*
一路 上
throughout the journey
def: LocationThru={route|道路}

In addition to denote the content sense which is the exact place where the subject cried, place adverb '*yilushang*' in example (13) also function a manner role to connect the subject 'she' and the verb 'cried'. To achieve function senses and content senses representations, function words are defined by its event role 'LocationThru', which is the abbreviation of 'Location Through' , and the content senses 'route' are defined in the value column after the event role. The definitions of other place adverbs are exemplified as follows.

(15) *dalaoyuan*
大老遠
from a distance
def: LocationIni={distant place|遠方}

(16) *bu yuan qianli*
不 遠 千里
from a distance
def: LocationIni={distant place|遠方}

(17) *sixia*
四下
all over the place
def: location={all|全}

(18) *yanjie*
沿街
along a street
def: LocationThru={route|道路}

(19) *daochu*
到處
everywhere
def: location={all|全}

(20) *yuci*
於此
here
def: position={here|這裡}

The semantic roles of Extended-HowNet form taxonomy. Most place adverbs play roles of location and its hyponymy. 'LocationIni' is the Initial Location where event occurred. 'LocationFin' is the final location toward which a motion action is directed. 'LocationThru' is the route or path through which an action is performed. Fig. 1 shows the taxonomy in Extended-HowNet.

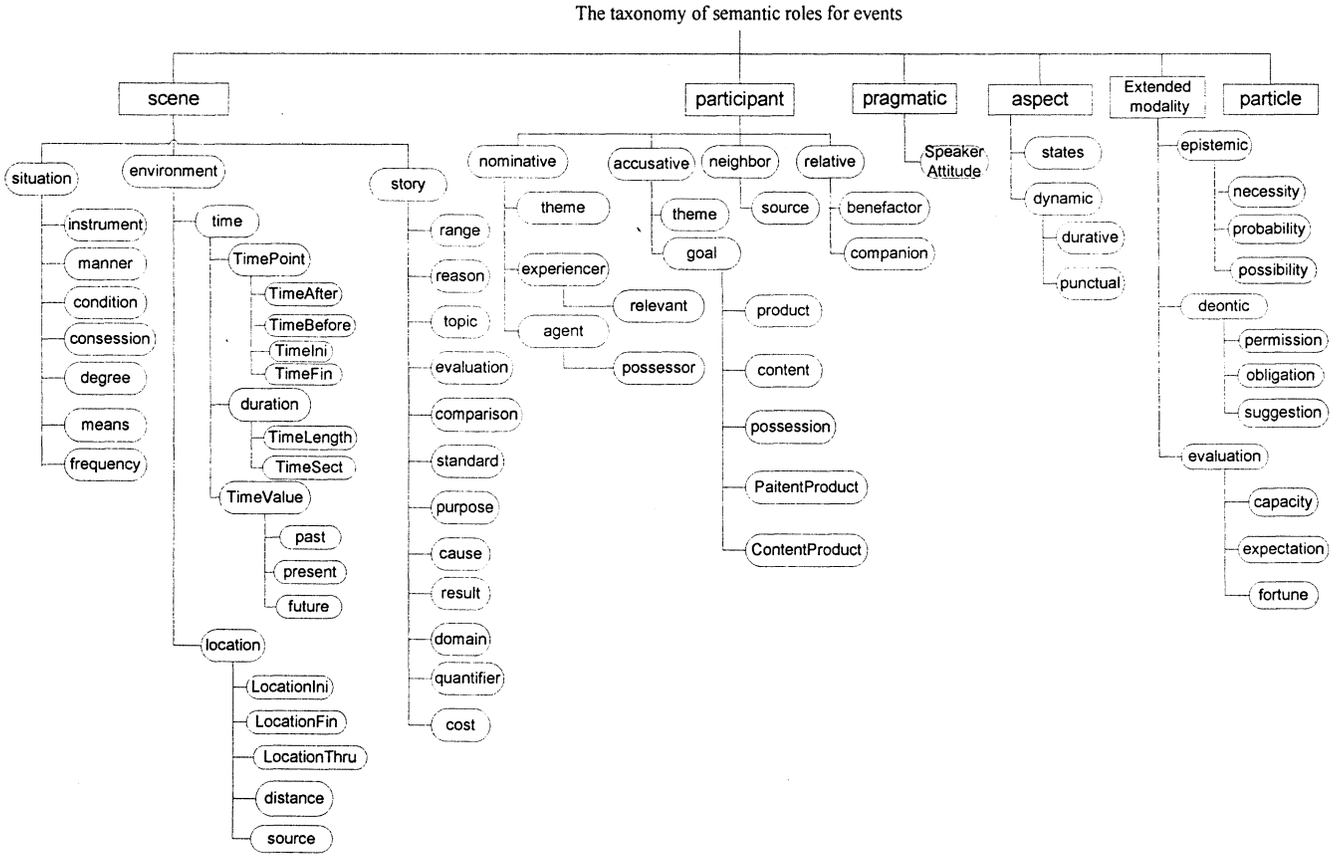

Fig. 1. The taxonomy of semantic roles for events.

However, there is a small part of place adverbs represent different semantic properties. Syntactically, they are all used to modify a verb, but some place adverbs' core meaning is referring to the subject itself, but not much focused on where the subject is connected with the event relatively. For example, in the sentence (21):

(21) *ta quanshen shitou*
他 全身 濕透
He soaks all over the body.

Although the place adverb 'all over the body' indicates the place where the event 'soak' happens, it's core meaning refers to the subject 'he', and indicates the subject is the very place that 'soak' happens. This is not a binary relation as the type we mentioned above any more, but it expresses a unary relation. So we define the place adverb 'all over the body' as (22):

(22) *quanshen*
全身
all over the body
def: location={whole(subject)}

We called this kind of definition a functional composition. It is used to describe the word sense which the argument of its semantic role doesn't point to its head, but point to the other part of the word sense. To show the difference of original definition and functional composition, we compared example (17) *sixia* with example (22) *quanshen*. '*Sixia*' is a typical example for original definition. For example, the sentence 'He looks all over the place.' can be represented as (23):

(23) *ta sixia zhangwang*
他 四下 張望
He looks all over the place.
def:{look|看:agent={he|他},target={all|全}}

In above definition, the arguments of both event roles 'agent' and 'target' are the head 'look', which normally be omitted while defining. On the contrary, when we describe '*quanshen*' in a sentence, the argument of the semantic role is not the head, but the other part of the sentence. For example, sentence (24) can be represented as:

(24) *ta quanshen shitou*
他 全身 濕透
He soaks all over the body.
def:{wet|濕: theme=whole(he|他)}

In above sentence, it is not going to show the precise meaning if we just define 'all over the body' as theme={all|全}, so we adopted the function feature *whole()* to indicate the argument of whole is the subject 'he', and represent it as whole(he) to show the unary relation exists between 'he' and 'all over the body'. This kind of expression is necessary and is quite different from HowNet's original mechanism which only expresses binary relations of words.

The semantic roles for functions in Extended-HowNet are shown in Fig. 2. The locational part of functions is what we focus on this paper.

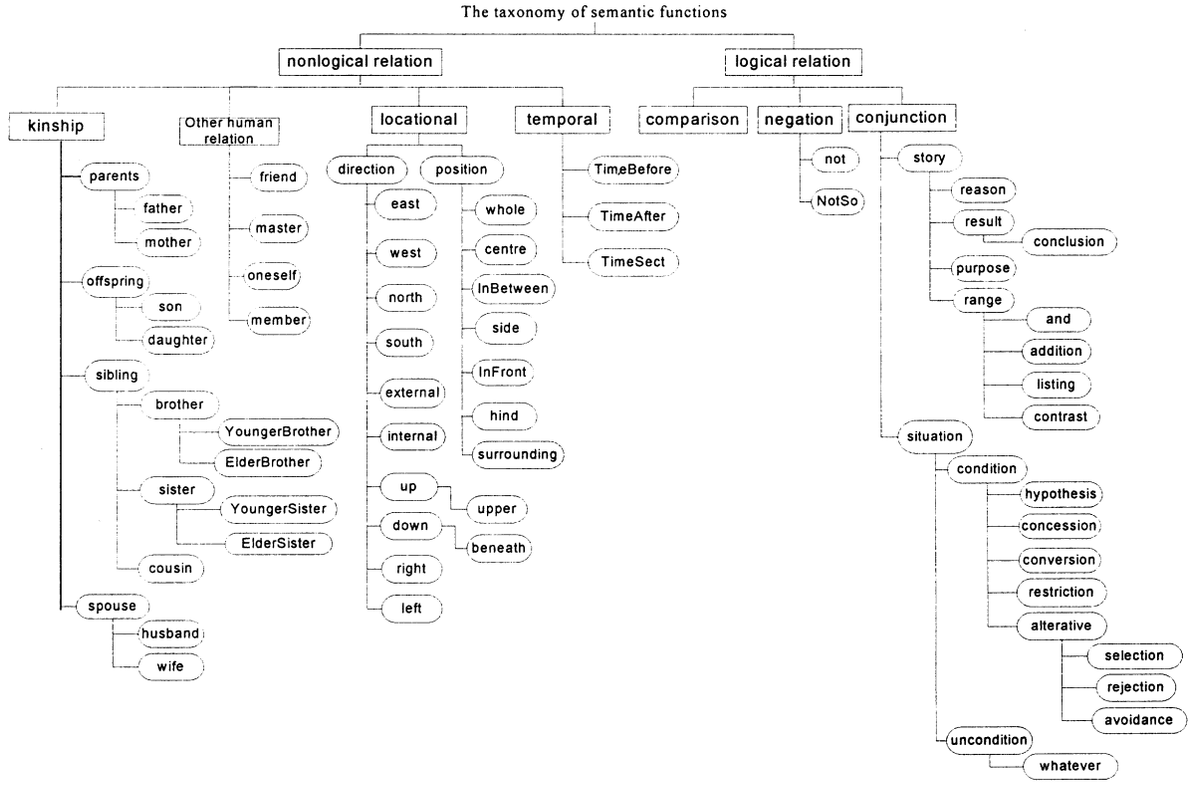

Fig. 2. The taxonomy of semantic functions.

C. *Place Prepositions*

A place preposition followed by a place noun and sometimes a localizer composing a locative phrase which play adverbial functions syntactically, such as *zai jia* 'at home', *zai louxia* 'downstairs', *zai jiejiao chu* 'round the corner' and *zai xiangzi li* 'in the box' etc. [7]. In Chinese, place preposition words are *zai*, *cong*, *dao*, *you,* and *zi* etc. Like other prepositions, all of place prepositions play a role to link a place to an event, so they are all binary relations. In addition, they own much less content senses, instead, their functional senses are strong. We defined them by representing representative event roles and leave their argument columns blank. The definitions are as follows:

(25) *zai*
在
at
def: location={}

(26) *cong*
從
from
def: LocationIni={}

(27) *dao*
到
to
def: LocationFin={}

(28) *you*
由
from
def: LocationIni={}

(29) *dacong*
打從
from
def: LocationIni={}

(30) *zi*
自
from
def: LocationIni={}

Finally, there are still some adverbs which contain spatial wording are not included in above discussion. It's worth discussing them to end this section. These words are *yingmian* 'face to face', *dangzhong* 'before the public', *suidi* 'at everywhere' and *shundao* 'by the way' etc. These adverbs describe binary relation in sentences. But if we use 'location' role to define them, it fails to express their precise meanings. On the contrary, 'manner' apparently is the very event role which can define them briefly and precisely. Therefore these words are defined as follows:

(31) *yingmian*
迎面
face to face
def: manner={facing|朝向}

(32) *dangzhong*
當眾
before the public
def: manner={overt|公開}

(33) *suidi*
隨地
at everywhere
def: manner={rash|莽}

(34) *shundao*
順道
by the way
def: manner={by the way|順便}

These examples tell us, the research of semantic knowledge representation might also provide us new idea about how to distinguish word senses.

IV. Semantic Composition For Spatial Concepts In Extended-Hownet

To analyze lexical semantic relation and defining senses is not only for distinguishing word senses, but considering the semantic composition from words to a phrase and phrases to a sentence. Since many spatial concepts own rich relational senses, they are good examples to demonstrate how word sense definitions can be unified into a sense representation of sentences in composing words into sentences.

Since both place adverbs and locative phrases function as modifiers and serve to express the locational relation in phrases or sentences, they should have the same or similar semantic representations after the semantic composition process. We can test this assumption in the following examples:

(35) *yanjie* *jiaomai*
沿街 叫賣
'peddle something in the street'
yanjie def:LocationThru={route|街道}
jiaomai def: {cry|喊:content={sell|賣},
domain={economy|經濟}}
Compositional Representation=
{cry|喊:
content={sell|賣},
domain={economy|經濟},
LocationThru={route|道路}}

(36) *yanzhe* *jiedao* *jiaomai*
沿著 街道 叫賣
'peddle something in the streets'
yanzhe def: LocationThru={}
jiedao def:{route|道路}
jiaomai def: {cry|喊:content={sell|賣},
domain={economy|經濟}}
Compositional Representation=
{cry|喊:
content={sell|賣},
domain={economy|經濟},
LocationThru={route|道路}}

(37) *jiadao* *huanying*
夾道 歡迎
'welcome around the road'
(or 'line the street to welcome')
jiadao def:location=side(route|道路)
huanying def:{welcome|歡迎}
Compositional Representation=
{welcome|歡迎:
location=side(route|道路)}

(38) *zai* *jiedao* *liangpang* *huanying*
在 街道 兩旁 歡迎
'welcome at both side of the road'
zai def: location={}
jiedao def:{route|道路}
liangpang def: location=side(x)
huanying def:{welcome|歡迎}
Compositional Representation=
{welcome|歡迎:
location=side(route|道路)}

(35) and (36) show different surface forms but identical deep semantics. After the semantic composition process, the Extended-HowNet representations come out to be the same for both examples. It is also the case for (37) and (38). The heartening result shows that the Extended-HowNet representations are near-canonical and worth further developing.

V. Conclusion And Future Work

We have shown a new lexical knowledge representation mechanism by demonstrating the different definition form of spatial concepts, include place adverbs, place nouns, and place prepositions. Differing from HowNet's knowledge representation mechanism, which only expressed the type of place to event relations of words, we proposed a functional mechanism which maps an argument from domain to its range. For instance, a directional function maps a place to another place. In defining spatial concepts, directional functions are mainly used to describe detailed direction or position. Spatial related words including few place adverbs, such as 'all over the body', and many place nouns, such as 'south east', 'the infield', 'river bank', 'inside the room' are defined by functions and functional compositions. The other part of spatial concepts, include the majority of place adverbs and all place prepositions, are defined by event roles which introduce to the location of an event.

In addition, since there is no clear-cut differentiation between content words and function words, a lexical knowledge representation system is necessary to encode both relational senses and content senses for each word. We adjust the HowNet definitions for function words to represent concepts more precisely and more reasonably such that meaning composition process can be carried out while composing phrases into a sentence.

A hierarchical structure of semantic roles is needed for feature unification in the process of semantic composition. In this article, we display the fine-grain semantic roles for spatial concepts. In the future, we will continue to study the detail representations for modality, negation, temporal relation, etc. The fine-grain semantic roles will also be future refinement of Extended-HowNet.

Acknowledgment

This research was supported in part by National Science Council under a Center Excellence Grant NSC93-2752-E-001-001-PAE This and Grant NSC93-2213-E-001-019.

References

[1] Z. Dong and Q. Dong, *HowNet*, http://www.keenage.com/
[2] C. Fellbaum, *WORDNET-An Electronic Lexical Database*, the MIT Press, 1998.
[3] T. Veale, "Analogy Generation with HowNet," *Proceedings of the International Joint Conference on Artificial Intelligence*, Edinburgh, 2005.
[4] K. Chen, S. Huang, Y. Shih, and Y. Chen, "Multi-level Definitions and Complex Relations in Extended-HowNet," Workshop on Chinese Lexical Semantics, Beijing University, 2004 (in Chinese).
[5] Y. Chen, S. Huang, Y. Shih, and K. Chen, "Semantic Representation and Definitions for Function Words in Extended-HowNet," Workshop on Chinese Lexical Semantics, Xiamen University, 2005 (in Chinese).
[6] CKIP (Chinese Knowledge Information Processing), *The Categorical Analysis of Chinese. CKIP Technical Report 93-05*, Nankang: Academia Sinica, 1993 (in Chinese).
[7] C. Li and S. Thompson, *Mandarin Chinese: a functional reference grammar.* University of California Press, Berkeley/Los Angeles, 1981.