A Unified Representation of Attributes and Their Semantic Composition

Su-Chu Lin, Wei-Yun Ma, and Yueh-Yin Shih
Institute Information Science
Academia Sinica
Taipei, Taiwan
e-mail: {jess, ma, yuehyin}@iis.sinica.edu.tw

Abstract—**Attributes are the way we describe the world. A simple classification on attributes' object host or event host is not enough for explaining appearance of both object attributes and event attributes. Thus we propose a presentation and explanation for attributes which introduce single and group host criterion for object attributes and linguistic valence for event attributes. Especially, we argue that all event attributes can be presented in a unified representation where an event attribute takes a certain latent event as its single host and this event's arguments may as well appear as the host arguments. In addition, we try to recover the omitted events in event attributes and express the semantic composition of the phrases in E-HowNet's framework.**

Keywords-attribute; E-HowNet; semantic expression; semantic composition

I. INTRODUCTION

Attributes, the criterion for classifying things, are the basis of understanding and distinguishing entities. For example, a table can be described by its material, size, shape and etc. and thus *material*, *size* and *shape* are some attributes of a table. We use these attributes to distinguish one object from another or to classify objects with the same criteria. Linguists define attributes as perceptive and cognitive aspects of objects acquired by human in the process of learning and constructing the world. In other words, attributes reflect our outlook on the world. However most famous ontologies, such as Sumo [1] or WordNet [2], focus more on nominal objects or predicative events and draw less attention on attributes. Fortunately E-HowNet [3] inherited a relatively elaborated taxonomy on attributes from HowNet [4]. With this advantage that E-HowNet brings, we are able to go deeper into linguistic phenomenon on attributes. Attributes are dependent from their hosts; that is, we need a specific host to understand the denotation of an attribute in context. This dependent relationship accounts for categorizing attributes by their hosts. Usually attributes are classified into object attributes or event attributes in the recent researches. But this classification does not explain attributes that may take both event and object hosts, such as in 偵查範圍 (*zhencha fanwei*, range of scouting) and 雷達範圍 (*leida fanwei*, range of the radar). This paper aims to propose a unified presentation and explanation for attributes that may have either or both object and event hosts. We especially focus on the analysis of the event attributes that take an event host, of which the object arguments will come naturally parts of the hosts. We argue that all event attributes can be presented in a unified representation. The basic form is an event attribute collocating with an event; the

transformation is an event attribute taking an object host, where the object is an argument of a certain latent event. We consider this latent event as a prerequisite in event attribute, even though this event might not be realized in the linguistic surface form. Information gap on this latent event is possibly filled up from context and common knowledge. Therefore the phrase 雷達範圍 (*leida fanwei*, range of the radar) should be recovered to 雷達偵查的範圍 (*leida zhencha de fanwei*, range of scouting by radar), where the event *scouting* is displayed as the nucleus of the event attribute *range*. This approach benefits in the unified representation of event attributes and makes it feasible in semantic composition. Moreover, the relations of object arguments revealed as agent, theme, instrument, etc. in event attributes are more fine-grained.

In addition to the introductory opening, this paper conveys previous works on attributes in the second section. The succeeding section introduces attributes in the framework of E-HowNet. In the third section we discussed on the classification of attributes. And followed by classification we display the syntactic and semantic expression of semantic composition. The finally section is our concluding remarks.

II. BACKGROUND

Most classifications on attributes are according to the criteria of their hosts [5] [6]. Liu [5] categorized attributes with the criteria of hosts and thus divided attributes into object attributes and event attributes. For object attributes, an object host often appears before them, consisting of an N-N phrase, e.g. 箱子的形狀 (*xiangzi de xingzhuang*, shape of the box). This is also an example of the inalienable possession between the attribute 形狀 (*xingzhuang*, shape) and the host 箱子(*xiangzi*, box). In contrast to the object attributes, event attributes take events as their hosts, constructing V-N compounds morphologically, e.g. 飛行速度 (*feixing sudu*. flying speed) and 創造能力 (*chuangzao nengli*, ability to create). Liu offered attribute categorizations and distinct linguistic observation on attributes and values, but explained nothing on the overlapping types.

With the similar foundation to Liu's classification, Xu [6] further classified attributes with four types of hosts, which are object, event, relation, and space, by using a feature distinction approach. Xu demonstrated attribute types with corresponding features and syntactic patterns. The relation host she provided clarifies the multiple objects host of the attributes and their particular syntactic behavior. This brings explanation to the distinct linguistic phenomenon. However the features she used to classify

event attribute, *[+active]* and *[-active]*, provide no account for event attributes taking two kinds of hosts. Event attributes demand more explanation than object attributes, due to the variety of collocation with their host modifiers. We intend to introduce linguistic valences for the clarification of attributes taking both object and event hosts.

The linguistic valences are the number of arguments controlled by predicates. Arguments are the required elements of the predicates, while adjuncts are optional expressions. Yuan [7] and Wang et al. [8] contributed to the analysis of the linguistic valence in Chinese, expanding the idea of verb valence to number of arguments dominated by nouns. The mono-valence nouns have dependent and inalienable possession with another object, such as the kinship noun 爸爸 (*baba*, father), the body part noun 尾巴 (*weiba*, tail) and the attribute noun 質地 (*zhidi*, texture). From the noun valence viewpoint, attributes are the mono-valence nouns. Moreover, Yuan [9] expounded on bi-valence nouns concerning with two entities, such as the word 意見 (*yijian*, opinion) in example (1).

(1) 老張對這件事的意見
Lauzhang dui zhejianshi de yijian
Lauzhang's opinion on this issue

The two valences of this bi-valence noun are 老張 (*Lauzhang*, a person named Lauzhang) and 這件事 (*zhejianshi*, this issue). Yuan supplemented another bi-valence noun type, the nominalized verbs, in a resembling construction as followed.

(2) 老張對這件事的決定
Lauzhang dui zhejianshi de jueding
Lauzhang's decision for this issue

Both nouns of 意見 (*yijian*, opinion) and 決定 (*jueding*, decision) dominate arguments of 老張 (*Lauzhang*, a person named Lauzhang) and 這件事 (*zhejianshi*, this issue) and thus they are bi-valence nouns.

III. ATTRIBUTES IN E-HOWNET

We remarked on the relatively elaborated taxonomy on attributes of E-HowNet in the introductory section. E-HowNet is an entity-relation model that represents lexical senses extended and evolved from HowNet [10] and it is further presented as an online common-sense-knowledge-based indexing relations of concepts obtained from lexicons [11]. In this framework, each concept is defined and associated with other concepts. Lexical sense definitions in E-HowNet provide rich information as well as encoded relational links between words via feature relations [12][13][14].

Attributes are an independent taxonomy in HowNet; on the other hand, E-HowNet view attributes as relation and subcategory of object, shown in Figure 1 [3]. This is in accordance with the usages of denoting relation in semantic description and serving as object nouns in linguistic realization.

Both relation and concept usage of attributes can be revealed in the semantic expression of E-HowNet format [15]. When attributes are concealed from words or phrases and only their corresponding hosts and values are clear, the relation expressions of attributes are therefore adopted. For example to express the sense of 快車 (*kuaiche*, fast train), the attribute *speed* is disclosed in the E-HowNet definition as {火車|train:speed={fast|快}}. It is used as a linkage between the host primitive *train* {火車|train} and the value primitive *fast* {fast|快}. And when attributes appear as heads in noun phrases like 步行速度 (*buxing sudu*, speed of walking), the semantic expression embraces an object-like definition in E-HowNet, shown in {speed|速度:host={walk|走}}. Note that in E-HowNet expression, the form of {speed|速度:host={walk|走}} is an equivalent conversion of {speed({walk|走})}, where the context is needed.

Therefore semantic representations of attributes in E-HowNet are either relations linking entities or concept entities as other object primitives. After reviewing attributes in E-HowNet, we will move on to the classification of attributes.
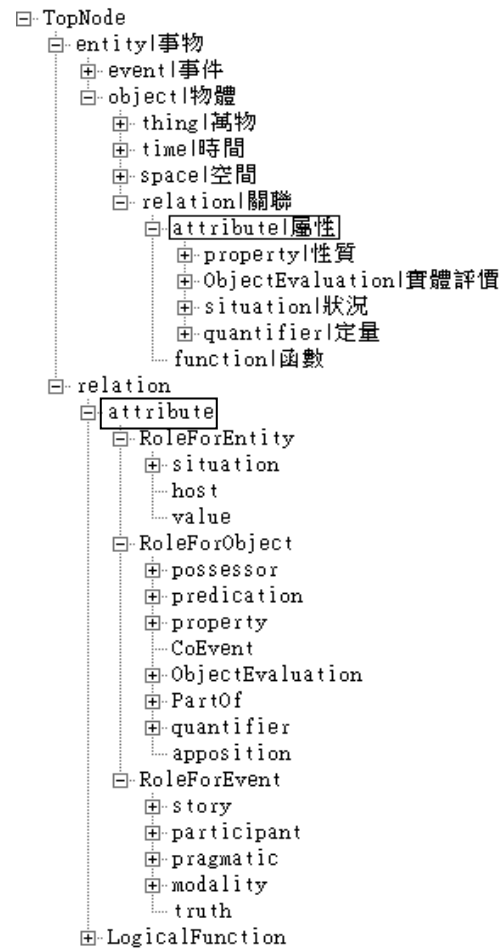

Figure 1. A clip of E-HowNet on tology on attributes

IV. ATRRIBUTE CATEGORIZATION

Following Liu's classification, we classify attributes on the criteria of their hosts, conducing to two types of attributes: one is object attribute and the other is event attribute. The two types of attributes can be sub-categorized by advanced host types.

A. Object Attributes

Object attributes describe properties of objects and often appear beforehand as a modifier. In the phrase 衣服尺寸 (*yifu chicun*, size of the clothes), the host 衣服 (*yifu*, clothes) modifies the object attribute 尺寸 (*chicun*, size) and its E-HowNet definition would be {size|尺寸:host={clothing|衣物})}. Similarly, the subject in the sentence of example (3) is composed with the semantic head attribute 人品(*renpin*, morality) and a human host 老張(Lauzhang), and its semantic expression is listed.

(3)老張的人品
Lauzhang de renpin
the morality of Lauzhang
{morality|道德:host={human|人:name={"老張"}}}

This type of attribute is fulfilled by a single host. On the other hand, Xu [6] suggests attributes taking two hosts or group host be categorized as relational attributes. The attribute noun 距離 (*juli*, distance) in example(4) takes a group host of 書本 (*shuben*, book) and 眼睛 (*yanjing*, eyes), connected by a conjunction 和 (*he*, and).

(4)書本和眼睛之間的距離
shuben he yanjing zhijian de juli
the distance between the table and the eyes
{distance|距離:host={InBetween({publications|書刊},{eye|眼})}}

When there is only one host appearing in the constituent with the attribute *distance*, like 書本的距離 (*shuben de juli*, distance of the books), we may interpret 書本 (*shuben*, book) as the non-inflectional plural noun, denoting as the distance between one book and another. Otherwise we expect another related host, such as 眼睛 (*yanjing*, eyes) to show up in the nearby context as shown in example (5). The particular syntactic behavior of group host attribute is self-evidently separated from single host attribute.

(5)書本的距離大概需要與眼睛保持三十公分以上
shuben de juli dagai xuyao yu yanjing baochi sanshi gongfen yishang
The distance between the book and the eyes probably should keep 30 centimeters or more.

From the discussion above, we conclude that the object attribute is divided according to the single host or group host it may take and become single host attribute or group host attribute, respectively.

B. Event Attributes

Unlike object attributes, event attributes require an event to fulfill their hosts. For example in the phrase 偵測範圍 (*zhencha fanwei*, range of scouting), we have the attribute *range* with an event *scouting* and it is semantically expressed as {range({scout|偵察})} in E-HowNet. Sometimes the same attribute *range* appears with an object host as *radar* in 雷達範圍 (*leida fanwei,* range of radar). One explanation is to classify the attribute *range* in both object and event attributes. However this approach encounters another classification problem when an object and an event host both show up with the attribute as in 雷達偵測範圍 (*leida zhencha de fanwei,* range of scouting by radar).

To solve the problem, we adopt linguistic valence to explain this phenomenon. First of all we claim that the event attributes demand an event no matter the event appear in the linguistic surface form or not. Therefore in the semantic definition of 雷達範圍 (*leida fanwei*, range of radar), we will introduce the omitted event *scouting* and the definition will be {range({scout|偵察:instrument={雷達|radar}})}. In this way we view radar as an argument of the omitted event. When an object host appears before an event attribute, the concealed event needs to emerge to fill the gap in nature language understanding. The valence of event attribute noun is concerned with the valence of event. Therefore valence type of event attribute is relating to the valence of the host event.

V. SEMANTIC COMPOSITION OF ATTRIBUTES AND HOSTS

We have discussed on the representation of attributes and their classifications. The complexity of event attributes appearing in a phrase has drawn our attention. We have discussed on recovering the omitted event in event attribute. It is easy for human but difficult for computers to fill up the gap of the omitted head verb in language. Fortunately there are still some clues for this inference. Take the word 雷達 (*leida*, radar) for example, it is a basic concept word {雷達|radar} in E-HowNet, and if we expand the expression to the next level, we will get def:{tool|用具:telic={scout|偵察:instrument={~}}}, which implies that radar is a {tool|用具} and this tool has a telic or function of {scout|偵察}. As a result, a candidate event *scouting* involved with radar is obtained. Another possible way to find the omitted event may come from the attribute itself. For example the event attribute 適應力 (*shiyingli*, adaptability) is supposed to contain an event *to fit* itself, and when there is an object appear as the host, we may infer the event *to fit* {fit|適合} in the E-HowNet expression as in example (6) and (7).

(6)幼兒的適應力
you'er de shiyingli
the adaptability for children
{ability|能力:host={fit|適合:theme= {小孩|child}}}

(7)對氣候的適應力
dui qihou de shiyingli
the adaptability on climate
{ability|能:host={fit|適合:content= {weather|天象}}}

In addition to the phrasal level, we may observe a similar phenomenon in the composition of sentence level. Since event attributes are able to take the events and their arguments in the host position, event attribute with multiple hosts is possibly seen in a sentence. For instance in (8), the event attribute is collocated with the event and two arguments of the event.

(8)雷達偵測敵機的範圍極廣

leida zhencha diji de fanwei ji guang

The range of radar scouting the enemy plane is very extensive.

The syntactic tree of the sentence above is shown in Figure 2, where the subject contains the event attribute *range* and the complete event arguments of the event *scouting*.
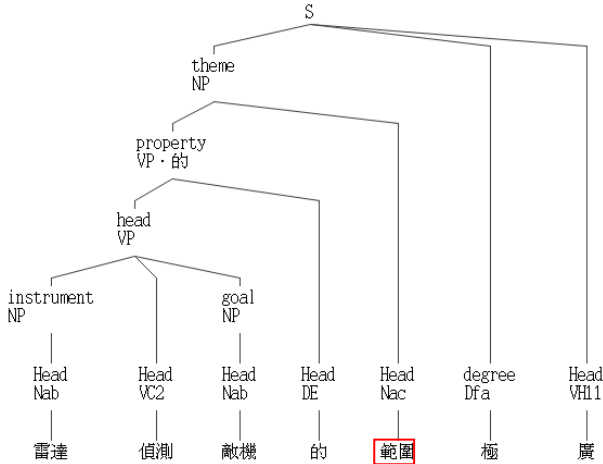

Figure 2.   Syntactic tree of the example sentence

The semantic definition of each word in example (8) is represented in upper columns of Figure 3 and we have semantic composition of each lexicon with syntactic brackets in the lower column. The semantic expression corresponds with the syntactic structure of this complex sentence.

Figure 3.   semantic representation of the example sentence

VI.   CONCLUSION

This paper proposes a unified presentation and explanation for attributes. We discuss on the attribute classification and semantic composition with their hosts. Followed the previous works, we classify attributes in the criteria of their hosts and thus group them into object attribute and event attributes. The object attributes are sub-divided into single host attributes and group host attributes, due to the distinct syntactic behavior of the group hosts. On the other hand, event attributes may take events and these events' object arguments as their hosts. The events are required in an event attribute and therefore they should be recovered even though sometimes are omitted in the linguistic surface form. As a result we propose some possible clues for the event recovering. Finally we demonstrate the semantic composition of attributes in E-HowNet expression. This representation not only makes it feasible in semantic composition but also provides more fine-grained relations of object arguments with events in event attributes.

In the future, we plan to develop a composer to automatically express the semantic composition of all attributes and their hosts in the framework of E-HowNet.

REFERENCES

[1]   SUMO, http://www.adampease.org/OP/

[2]   WordNet, https://wordnet.princeton.edu/

[3]   E-HowNet, http://ehownet.iis.sinica.edu.tw/

[4]   Z. Dong, and Q. Dong, HowNet, http://www.keenage.com/

[5]   C. Liu, "Studies on Attribute Domain in Contemporary Chinese", Sichuan: SCPG, 2008.

[6]   Y. Xu, "A Study in the Semantic Feature of Attribute Nouns in Modern Chinese", Wuhan: Wuhan University Press, 2013.

[7]   Y. Yuan, "A Cognitive Study of the Mono-valent Noun in Mandarin Chinese", in Studies of the Chinese Language, Peking: Institute of Linguistics Chinese Academic of Social Sciences, 1994, pp. 241-253.

[8]   H. Wang, W. Zhang and S. Yu, "Structure and Application of The Semantic Knowledge base of Contemporary Chinese," in Applied Linguistics, vol. I, 2006, pp.134 141.

[9]   Y. Yuan, "Bi-valent Nouns in Modern Chinese," in Studies on valent grammar, Y. Shen and D. Zheng, Eds. Peking: Peking University Press, 1995, pp.29-58.

[10]   Z. Dong, and Q. Dong, HowNet and the Computation of Meaning, World Scientific Publishing Co. Pte. Ltd, 2006..

[11]   S. Huang, Y. Chung and K. Chen, "E-HowNet: the Expansion of HowNet", the First National HowNet Workshop, Beijing, China, 2008.

[12]   K. Chen, S. Huang, Y. Shih, and Y. Chen, "Extended-HowNet- A Representational Framework for Concepts," OntoLex 2005-Ontologies and Lexical Resources IJCNLP-05 Workshop, Jeju Island, South Korea, 2005.

[13]   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.

[14]   CKIP (Chinese Knowledge Information Processing Group), E-HowNet Technical Report, CKIP, IIS, Academia Sinica, Taipei 2009.

[15]   S. Lin, S. Huang, Y. Chung, and K.Chen, "The Lexical Knowledge and Semantic Representation of E-HowNet", Contemporary Linguistics, vol.15, no.2, 2013, pp.177-194