

Knowledge Representation and Sense Disambiguation for Interrogatives in E-HowNet

Shu-Ling Huang^{**} and Keh-Jiann Chen^{*}

Abstract

In order to train machines to ‘understand’ natural language, we propose a meaning representation mechanism called E-HowNet to encode lexical senses. In this paper, we take interrogatives as examples to demonstrate the mechanisms of semantic representation and composition of interrogative constructions under the framework of E-HowNet. We classify the interrogative words into five classes according to their query types, and represent each type of interrogatives with fine-grained features and operators. The process of semantic composition and the difficulties of representation, such as word sense disambiguation, are addressed. Finally, machine understanding is tested by showing how machines derive the same deep semantic structure for synonymous sentences with different surface structures.

Keywords: Semantic Representation, Sense Disambiguation, Interrogatives, E-HowNet

1. Introduction

Electronic dictionaries are designed for the purpose of providing users (or computers) convenient access to relevant knowledge of words to understand language. When we say that a sentence is ‘understood’, we mean that the concepts and the conceptual relations expressed by the sentence are unambiguously identified and we can make the correct inferences/responses. To have a computer understand a sentence, we must have a framework for representing lexical knowledge and performing semantic composition and disambiguation processes.

Extended-HowNet (E-HowNet for short) is a frame-based entity-relation knowledge representation model, which was extended from HowNet [Dong *et al.* 1988] to encode concepts. Concepts are represented and understood by their definitions and association links to

*CKIP, Institute of Information Science, Academia Sinica

E-mail: kchen@iis.sinica.edu.tw

[†]Department of Language and Literature Studies, National Hsinchu University of Education

E-mail: slhuang@mail.nhcue.edu.tw

other concepts. Compared to WordNet, HowNet's architecture provides richer information apart from hyponymy relations. It also enriches relational links between words via encoded feature relations. The advantages of HowNet are (a) its inherent properties are derived from encoded feature relations in addition to hypernym concepts, and (b) information regarding conceptual differences between different concepts and information regarding morph-semantic structure are encoded. Therefore, we adopt a similar mechanism to define word sense in E-HowNet, but represent concepts in a more accurate and flexible way by (a) defining new concepts by well-defined concepts, (b) providing a uniform representational framework for both function words and content words, and (c) embedding semantic composition and decomposition capabilities. More detailed discussions can be seen at [Chen *et al.* 2005].

In E-HowNet, we define each lexical sense by the composition of well-defined concepts and/or basic concepts, called sememes in HowNet. The sememes are linked to their sense equivalence WordNet synsets [Fellbaum 1998]. Take 土地公 'God of earth' as an example:

- (1) 土地公 *Tu di gong* 'God of earth'
 Def: {God|神:telic={manage|管理:patient={land|陸地},agent={~}}}

Here, 'God' is the hypernym of the target word 土地公 'God of earth', 'manage' and 'land' are its related concepts. 'telic', 'patient', and 'agent' are relations which link these concepts. Obviously, to achieve mechanical understanding of natural language, the same or similar concepts must have the same or similar underlying semantic representation. However, natural language can be ambiguous. Different sentences might express the same meaning, and the same sentence can also express different meanings. The following sentences (2) and (3) show the former phenomenon, and (4) and (5) show the latter:

- (2) 我能否拍照? *Wo neng fou pai zhao?* Is it OK for me to take pictures?
 (3) 我可不可以照相? *Wo ke bu ke yi zhao xiang?* Can I take photos?
 (4) 土地公有政策。 *Tu di gong you zheng ce.* The policy of public sharing of the land.
 (5) 土地公有政策。 *Tu di gong you zheng ce.* God of earth has his policy.

Thus, transforming the surface structure of a sentence into a canonical semantic representation

and simultaneously solving the problem of word sense ambiguity are major research issues. In summary, lexical semantic representation and composition (including disambiguation) are the most demanding techniques for understanding natural language by machines and the design of E-HowNet is aimed for these objectives.

In this paper, we will take interrogatives as examples to demonstrate the mechanism of lexical semantic representation and composition in E-HowNet [Chen *et al.* 2004]. The goal is to achieve near canonical semantic representation for synonyms and sense equivalent sentences. Take sentences (2) and (3) as examples. Although their syntactic structure and surface strings are very different, by composing lexical sense representations, we hope the machine can ‘understand’ synonymy of sentences in different surface forms.

Analysis of interrogative constructions is of great interest to linguists, as well as to computer scientists, for example, those who are engaged in QA techniques. Interrogative constructions have played a central role in the development of modern syntactic theory. Ginzburg and A. Sag [2000] have pointed out that the interrogative has been at the heart of work in generative grammar, along with government and binding (GB) theory and head-driven phrase structure grammar (HPSG). Nonetheless, to date, most syntacticians take quite different approaches from semantic and pragmatic points of view on interrogatives. Taking questions in Mandarin Chinese as example, Shao [1996] has summed up the current study of interrogatives and listed the main research themes as follows: the type of question, interrogative particles, querying focus and its answer, degree of doubt, special interrogative sentence patterns, etc. Most of the above themes are purely grammatical analysis. To build a frame-based entity-relation knowledge representation model, we find interrogative construction a good and challenging example because it combines problems of syntax, semantics, and pragmatics.

In the following section, we briefly describe the previous works for interrogatives. Then, we introduce our analysis of type classifications for interrogatives and their representation in E-HowNet. Next, we present the semantic composition process of interrogative sentences and the difficulties encountered. We conclude the paper by discussing our results and future work.

2. Background

Interrogatives in Chinese studies are traditionally attributed to the mood category of syntax. Ma [1935] wrote the first grammar book for Mandarin Chinese. He classified interrogatives into the mood category. Later, Li [1930] and Lv [1942] carried forward his viewpoint and influenced modern linguistic theory on interrogatives deeply. Most linguists consider there to be four grammatical types that explicitly mark an utterance as an interrogative [Lv 1942; Li and Thompson 1997; Tang 1983; Lu 1984; Shao 1996]. First, is a question which can be answered by ‘yes’ or ‘no’, called a factual question, true and false interrogative, or yes/no

- (8) interrogative –(A) true/false (yes/no) interrogative
 - (B) Wh-interrogative – (a) asking factual information
 - (b) asking relationship
 - (c) asking opinion
 - (d) option choosing

Due to the different querying focus, we separate (A) true/false interrogative from (B) Wh-interrogative. Take sentences (9), (10) as examples:

- (9) 你喜不喜歡這個遊戲?
Ni xi bu xi huan zhe ge you xi?
Do you like this game or not?

- (10) 有誰知道我可以在哪裡找到這個遊戲?
You shei zhi dao wo ke yi zai na li zhao dao zhe ge you xi?
Who knows where I can find this game?

Sentence (9) belongs to the true/false interrogative type and the entire statement is a querying focus. Dissimilarly, Sentence (10) indicates two querying foci through using different interrogatives, ‘who’ and ‘where’. In other words, the true/false interrogative asks truth value of the positive predication of the sentence. Then, the Wh-interrogative is used to ask for information. By analyzing the querying focus, Wh-interrogatives can be further divided into four types: (B-a) asking factual information, such as time, location, quantity, and so forth; (B-b) asking relationship, such as kinship; (B-c) asking opinion or attitude, such as possibility, capacity, volition, etc.; and the last, (B-d) asking to choose an option. Sentence (10) refers to type (B-a). For the remaining types, we give an example of each as follows:

- (11) 她是你的什麼人?
Ta shi ni de she me ren?
What is the relationship between you and her?

- (12) 他可不可以吃辣椒?
Ta ke bu ke yi chi la jiao?
Can he eat hot peppers?

- (13) 淘米水是酸性還是鹼性？

Tao mi shui shi suan xing hai shi jian xing?

Rice rinsing water is acidic or alkaline?

Here, the fine-grain distinctions between interrogative type (B-a) and (B-b) and between interrogative type (A) and (B-c) are clarified below. For instance, sentence (11) refers to type (B-b), but why do we need to separate it from type (B-a) when they both use ‘what’ to make questions? In sentence (11), the question word 什麼 *she me* ‘what’ asks for relationship but not the type of a frame element or the value of a semantic role as exemplified in (10).² The semantic representation of a complex relation is different from the representation of entities. Therefore, we differentiate between interrogative type (B-a) and (B-b). Chen *et al.* [2004] proposed a compositional mechanism to describe complex relations. For example, we express ‘mother in law’ as (14):

- (14) mother in law

def:{human|人=mother(spouse({x:human|人}))}

According to the representation model, when our querying focus is a complex relation, we put a question mark before the relation role, such as mother, spouse, parents, etc., to mark the query focus. Representational detail is shown in the next section.

Second, some may argue that there is no distinction between type (A) and (B-c). Example (12) is of type (B-c); we find it may have a yes/no answer, a typical characteristic of type (A).

- (12) 他可不可以吃辣椒？

Ta ke bu ke yi chi la jiao?

Can he eat hot peppers?

However, if we compare the meaning of sentences (12) and (15), we can still find a slight difference between a yes/no question and a question of asking opinion. Sentence (12) has the meaning of asking the hearer’s permission, but (15) does not.

² The disambiguation of 什麼, see section 4.1.1.

(15) 他喜不喜歡吃辣椒？

Ta xi bu xi huan chi la jiao?

Does he like to eat hot peppers?

Therefore, in Mandarin Chinese, 可不可以 *ke bu ke yi* ‘can or cannot’ and 喜不喜歡 *xi bu xi huan* ‘like or do not like’ both have the *A not A* form, but from the semantic point of view, they belong to different types. Shao [1996] has classified the *A not A* form into five classes according to *A*’s part of speech, shown as follows: (a) *A* is a copula *e.g.* 是不是 *shi bu shi* ‘be or be not’ (b) *A* is a modal word *e.g.* 好不好 *hao bu hao* ‘ok or not ok’ (c) *A* is an auxiliary *e.g.* 肯不肯 *ken bu ken* ‘willing or not willing’ (d) *A* is a verb *e.g.* 懂不懂 *dong bu dong* ‘understand or not understand’ (e) *A* is an adjective *e.g.* 美不美 *mei bu mei* ‘beautiful or not beautiful’. From the semantic perspective, we merge (a), (d), (e) and (b), (c) to re-divide these five categories into two categories, *i.e.* modal *A not A* interrogatives and other *A not A* interrogatives.

3.2 Knowledge Representation for Interrogatives

In E-HowNet, we made distinctions between content sense and relational sense thereby representing senses of content words and senses of function words in different ways. For instance, in (16), the content words ‘bathe’ and ‘cold water’ are represented differently from the function word ‘with’. In this example, the function word ‘with’ plays the role of ‘instrument’ which links the relation between its argument ‘cold water’ and the matrix verb ‘bathe’.

(16) bathe def: {clean|使淨: patient={body|身體}}
with def: instrument={}
cold water def: {water|水: temperature={cold|冷}}.

Result of semantic composition:

{clean|使淨: patient={body|身體}, instrument={water|水:temperature={cold|冷}}}.

In much the same way, interrogative words have more relational sense than content sense, so they are defined by semantic role to denote relational sense and use the operator ‘.Ques.’ to mark the querying focus, *i.e.* the object or its discrimination features which speakers want to know.

According to the classification of interrogatives above, we represent each type of interrogative as follows:

- (17) true/false interrogative def: truth={.Ques.}
- Wh-word interrogatives
- asking factual information def: role={.Ques.}
- asking relationship def: .Ques.RelationRole()
- asking opinion def: ModalityRole={.Ques.}
- option choosing def: role={.Option.{{x}.or.{y}}}

We use two operators, .Ques. and .Option., to denote querying focus or optional items. More detailed sense representations for interrogatives are shown in Table 1.

Table 1. The Type Classification and Semantic Representation of Interrogatives

Question types	Words	Sense representation
true/false interrogatives	嗎 <i>ma</i> ³ 呢 <i>ni</i> 是否 <i>shi fou</i> ‘whether’ 有沒有 <i>you mei you</i> ‘have’ 是不是 <i>shi bu shi</i> ‘is it’ 不是嗎 <i>bu shi ma</i> ‘isn’t it’ <i>A not A</i>	def: truth={.Ques.}
Wh-word interrogatives: asking factual information	誰 <i>shei</i> ‘who’	def: participant={animate: formal={.Ques.}}
	幾點鐘 <i>ji dian zhong</i> ‘what time’	def: time={.Ques.}
	什麼 ^① <i>she me</i> ‘what’ 何 ^① <i>he</i> ‘what’ 啥 ^① <i>sha</i> ‘what’	def: participant={inanimate: formal={.Ques.}}
	什麼 ^② <i>she me</i> ‘what’ 何 ^② <i>he</i> ‘what’ 啥 ^② <i>sha</i> ‘what’	def: formal={.Ques.}
	爲何 <i>wei he</i> ‘why’ 何以 <i>he yi</i> ‘why’ 何不 <i>he bu</i> ‘why not’	def: reason={.Ques.}

³ In this paper, our focus is semantic representation, so we don’t discuss the interrogative words ‘啊 *a*’; ‘呀 *a*’; ‘囉 *luo*’; ‘乎 *hu*’; ‘吧 *ba*’ or ‘喔 *wo*’. This is because the tone decides if they are interrogative words or not.

	爲什麼 <i>wei she me</i> 'why' 幹嘛 <i>gan ma</i> 'why' 怎麼 ^① <i>ze me</i> 'why'	
	哪裡 <i>na li</i> 'where' 哪兒 <i>na er</i> 'where' 哪 ^① <i>na</i> 'where'	def: location={.Ques.}
	哪 ^② <i>na</i> 'which' 哪些 <i>na xie</i> 'which'	def:quantifier={.Ques.}(,quantity={mass 眾})
	怎麼 ^① <i>ze me</i> 'how' 怎(麼)樣 ^① <i>ze (me) yang</i> 'how to' 如何 ^① <i>ru he</i> 'how to'	def: means={.Ques.}
	怎(麼)樣 ^② <i>ze (me) yang</i> 'how come' 如何 ^② <i>ru he</i> 'how come'	def: participant={.Ques.}
	多 <i>duo</i> 'how'	def: degree={.Ques.}
	多少 <i>duo shao</i> 'how much' 幾 ^① <i>ji</i> 'how much'	def: quantity={.Ques.}
	幾 ^② <i>ji</i> 'how much'	def: ordinal={.Ques.}
asking relationship	什麼關係 <i>she me guan xi</i> 'what relationship' X 的什麼人 <i>x de she me ren</i> 'what relationship'	def: .Ques.RelationRole()
asking opinion	可不可以 <i>ke bu ke yi</i> 'can'	def: allowance={.Ques.}
	好不好 <i>hao bu hao</i> 'is it ok'	def: willingness={.Ques.}
	能不能 <i>neng bu neng</i> 'be able to' 能否 <i>neng fou</i> 'be able to'	def: capacity = {.Ques. }
	莫非 <i>mo fei</i> 'are' 難道 <i>nan dao</i> 'are'	def: possibility = {.Ques. }
choosing options	還是 <i>hai shi</i> 'or' 或 <i>huo</i> 'or'	def: role={.Option.{{x}.or.{y}}}

The interrogatives above are gathered from Li and Thompson's analysis and are integrated by checking over 1000 question titles manually in *Baidu knows* [Li *et al.* 1999]. To check the completeness of the above table and to find the distribution of query types, we randomly extracted 1% of the sentences with question marks from the Sinica corpus to see the coverage of the above table, and the results of the distribution are shown in Table 2. There are 203 sentences, 9 of which do not contain any query word listed in Table 1. Their query sense is expressed only by a question mark in the end of the sentence. However, we can insert 'ma' 嗎 or 'ne' 呢 before the question mark to these sentences to make them true/false interrogatives, so statistically they are still counted as true/false interrogatives. We can conclude that most of the questions ask true/false value or factual information. Furthermore, we can classify each of them according to Table 1.

Table 2. The Distribution of Question Type

Question types	Words	Number	Total(203)
true/false interrogatives	嗎 <i>ma</i>	50	98
	呢 <i>ni</i>	16	
	是否 <i>shi fou</i> 'whether'	10	
	有沒有 <i>you mei you</i> 'have'	1	
	是不是 <i>shi bu shi</i> 'is it'	6	
	不是嗎 <i>bu shi ma</i> 'isn't it'	1	
	<i>A not A</i>	14	
Wh-word interrogatives: asking factual information	誰 <i>shei</i> 'who'	4	94
	幾點鐘 <i>ji dian zhong</i> 'what time'	2	
	什麼 ^① <i>she me</i> 'what'	14	
	何 ^① <i>he</i> 'what'	1	
	啥 ^① <i>sha</i> 'what'	0	
	什麼 ^② <i>she me</i> 'what'	13	
	何 ^② <i>he</i> 'what'	2	
	啥 ^② <i>sha</i> 'what'	0	
	爲何 <i>wei he</i> 'why'	2	
	何以 <i>he yi</i> 'why'	0	
	何不 <i>he bu</i> 'why not'	1	

	為什麼 <i>wei she me</i> ‘why’	9	
	幹嘛 <i>gan ma</i> ‘why’	1	
	怎麼 ^① <i>ze me</i> ‘why’	3	
	哪裡 <i>na li</i> ‘where’	4	
	哪兒 <i>na er</i> ‘where’	1	
	哪 ^① <i>na</i> ‘where’	0	
	哪 ^② <i>na</i> ‘which’	5	
	哪些 <i>na xie</i> ‘which’	3	
	怎麼 ^③ <i>ze me</i> ‘how’	5	
	怎(麼)樣 ^④ <i>ze (me) yang</i> ‘how to’	3	
	如何 ^① <i>ru he</i> ‘how to’	8	
	怎(麼)樣 ^⑤ <i>ze (me) yang</i> ‘how come’	4	
	如何 ^② <i>ru he</i> ‘how come’	1	
	多 <i>duo</i> ‘how’	3	
	多少 <i>duo shao</i> ‘how much’	2	
	幾 ^① <i>ji</i> ‘how much’	3	
	幾 ^② <i>ji</i> ‘how much’	0	
asking relationship	什麼關係 <i>she me guan xi</i> ‘what relationship’	0	0
	X 的什麼人 <i>x de she me ren</i> ‘what relationship’	0	
asking opinion	可不可以 <i>ke bu ke yi</i> ‘can’	2	10
	好不好 <i>hao bu hao</i> ‘is it ok’	1	
	能不能 <i>neng bu neng</i> ‘be able to’	4	
	能否 <i>neng fou</i> ‘be able to’	2	
	莫非 <i>mo fei</i> ‘are’	0	
	難道 <i>nan dao</i> ‘are’	1	
choosing options	還是 <i>hai shi</i> ‘or’	1	1
	或 <i>huo</i> ‘or’	0	

3.3 Knowledge Representation for Interrogative Compounds

In addition to the above interrogatives, there are also some derived interrogative compounds, which are impossible to list comprehensively in our database. Thus, generating their sense representation automatically has to be accomplished. Most interrogative compounds are composed of an interrogative determinative and a noun. Since the number of interrogative determinatives is limited, have already been defined in Table 1, and listed in (18), we design the rules in (19) to derive the sense representations of interrogative compounds.

- (18) The sense representation of interrogative determinatives (ID)

什麼 <i>she me</i> /何 <i>he</i> / 啥 <i>sha</i> ‘what’	def:formal={.Ques.}
多少 <i>duo shao</i> ‘how many’	def:quantity={.Ques.}
幾 <i>ji</i> ‘how many’ or ‘what’	def:quantity={.Ques.} or def:ordinal={.Ques.}
哪 <i>na</i> ‘which’ or ‘where’	def:quantifier={.Ques.} or def:location={.Ques.}

- (19) The rules for deriving sense representations of interrogative compounds

If the morphological structure of the interrogative compound is **ID+head**.

- (A) **if head !=semantic role then def: {head:ID}**

e.g. 何人 *he ren* ‘who’ where ID= ‘何’ and head= ‘人’ ; hence after composition process we derive 何人 =def: {human|人:formal={.Ques.}}. Similarly,

e.g. 幾人 *ji ren* ‘how many people’ =def: {human|人:quantity={.Ques.}}

- (B) **if head=semantic role then def: role={.Ques.}**

e.g. 何時 *he shi* ‘when’ =def: time={.Ques.}

e.g. 什麼形狀 *she me xing zhuang* ‘what shape’ =def: shape={.Ques.}

e.g. 哪色 *na se* ‘what color’ =def: color={.Ques.}}

- (C) **if head=DM(dummy measure words) then def: ID**

e.g. 幾次 *ji ci* ‘how many times’ =def: quantity={.Ques.}

e.g. 哪隻 *na zhi* ‘which one’ =def: quantifier={.Ques.}

- (D) **if head=MM(meaningful measure words)⁴ then def: {MM: {ID} }**
e.g. 多少碗 *duo shao wan* ‘how many bowls’ =def:container={bowl|碗:
quantity={.Ques.}}

Other rules,

- (E) **while {第 *di*, 禮拜 *li bai*, 星期 *xing qi*, 西元 *xi yuan*, 民國 *ming quo*}+幾 *ji*
or 幾 *ji* + {點(鐘) *dain(zhong)*, 月 *yue*, 號 *hao*, 歲 *sui*, 週年 *zhou nian*, 年 *nian*}**
then define 幾 *ji* as def: ordinal={.Ques.}
otherwise define 幾 *ji* as def: quantity={.Ques.}
e.g. 第幾次 *di ji ci* ‘which time’ according to C and E def:
ordinal={.Ques.}
e.g. 幾點鐘 *ji dian zhong* ‘what hour is it’
according to D and E def: time={hour|小時: ordinal={.Ques.}}

- (F) **while verb+哪 *na* then define 哪 *na* as def: location={.Ques.}**
otherwise define 哪 *na* as def: quantifier={.Ques.}
e.g. 去哪 *qu na* ‘go where’ def: {go|去: location={.Ques.}}
e.g. 丟哪 *diu na* ‘throw where’ def: {throw|扔: location={.Ques.}}
e.g. 哪天 *na tian* ‘which day’ def: {day|日: quantifier={.Ques.}}

As the question determinatives 幾 *ji* and 哪 *na* are ambiguous, in addition to the general rule (A),(B),(C),(D), we also provide two other rules (E) and (F) to disambiguate their word sense.

4. Semantic Composition for Interrogatives

The previous discussion has been about semantic representation of lexical senses. To establish a formal system to handle the task of language understanding, we also need to address the issue of semantic composition. To understand Chinese sentences, after the word segmentation and parsing process, we get coarse-grained head-argument event structures of the sentences. Take Sentence (20) as an example:

⁴ The meaningful or dummy measure words are listed in Tai C.H. *et al.* “A Semantic Composition Method for Deriving Sense Representations of Determinative-Measure Compounds in E-HowNet” in the proceeding of *ROCLING 2008*, Taipei, Taiwan.

(20) 資料因何漏失？

Zi liao yin he lou shi?

Why is the data missing?

The segmentation and parsing result of (20) is:

Theme[NP:資料 *data*]+ reason[Dj:因何 *why*]+Head[VJ3:漏失 *lose*]

Then, we try to project surface syntax onto the semantic structure for establishing truly integrated semantic relations. In Example (20), we identify the sentential head ‘lose’ after the parsing process, and, based on E-HowNet, the arguments of event ‘lose’ are ‘possessor’ and ‘possession’; thus we know the ‘data’ here is the possession of ‘lose’. Therefore, the result of composition is as follows:

def: {lose|失去:possession={information|訊息},reason={ .Ques. }}

The other types of interrogative structures are exemplified as follows:

true/false interrogative

(21) 他病了嗎？

Ta bing le ma?

Is he sick?

def: {sick|病:experiencer={he|他},truth={ .Ques. }}

(22) 你是否曾說過謊？

Ni shi fou ceng shuo guo huang?

Have you ever lied before?

def: {lie|說謊:agent={listener|聽者},time={past|過去},truth={ .Ques. }}

Wh-word interrogative:

asking factual information

(23) 衣服上的墨水怎麼洗掉？

Yi fu shang de mo shui ze me xi diao?

How to wash away ink stains on cloth?

def: {wash|洗掉:patient={ink|墨水:place={clothes|衣服}},means={ .Ques. }}

(24) 哪些桌子壞掉？

Na xie zhuo zi huai diao?

Which tables are broken?

def: {OutOfOrder|壞:theme={table|桌子:quantifier={ .Ques. },quantity={mass|眾}}}

asking relationship

(25) 她是你的什麼人？

Ta shi ni de she me ren?

What is the relationship between you and her?

def: {he|他=.Ques.kinship(listener|聽者)}

asking opinion

(26) 莫非這是鬼城？

Mo fei zhe shi qui cheng?

Is it possible a ghost town?

def: {be|是:relevant={this|這},content={ghost town|鬼城},possibility={ .Ques. }}

(27) 由你開車行不行？

You ni kai che xing bu xing?

Is it ok for you to drive?

def: {drive|開車:experiencer={listener|聽者},willingness={ .Ques. }}

choosing options

(28) 他在這兒還是那兒住？

Ta zai zhe er hai shi na er zhu?

Does he live here or there?

def: {live|住:agent={he|他},location={ .Option. { {here|這兒}.or. {there|那兒} }}

(29) 他跪下來還是站在那裡求張三？

Ta qui xia lai hai shi zhan zai na li qiu zhang san?

Does he kneel on the ground or stand there to beg ChangShan?

def: {beg|求:agent={be|他},target={ChangShan|張三},means={ .Option. { {stand|站}.or. {kneel|跪} }}

Sense Disambiguation

To achieve the goal of automatic semantic composition, we have to solve the problem of word sense disambiguation. In Chinese, 什麼 *she me*, 怎麼 *ze me*, 怎樣 *ze yang*, 多 *duo*, 哪 *na*, and 幾 *ji* are the most frequently used interrogatives, and they all have ambiguous senses. In addition to Rule (19), their sense disambiguation rules are discussed below:

什麼 *she me* /何 *he*/ 啥 *sha*

什麼 *she me* ‘what’ plays the grammatical functions of adjective and pronoun, and there are two senses for each function. Accordingly, we generate four rules to disambiguate the word senses of 什麼 *she me*, and the details are shown in Table 3. 何 *he* and 啥 *sha* are its literary and slangy usage, hence, share the same disambiguation rules.

Table 3. Disambiguation Rules for ‘什麼 *she me*’

Rules of disambiguation	Examples & E-HowNet representation
adjective 1: Case 1: if 什麼/何/啥+semantic role then 什麼/何/啥=def: role={.Ques.} (什麼 ask the value of the semantic role.)	什麼時間 <i>she me shi jian</i> ‘what time’ def: time={.Ques.} 啥價錢 <i>sha jia qian</i> ‘what price’ def: cost={.Ques.} 何地 <i>he di</i> ‘what place’ def: location={.Ques.}
adjective 2: Case 2: if 什麼/何/啥+entity (nominalized verbs are included) then 什麼/何/啥=def: {entity:formal={.Ques.}} (什麼 ask the type/restriction of a frame element/participant role.)	何人 <i>he ren</i> ‘what person’ def: {human 人: formal={.Ques.}} 啥變化 <i>sha bian hua</i> ‘what change’ def: {change 變化: formal={.Ques.}} 什麼不同 <i>she me bu tong</i> ‘what difference’ def: {difference 不同: formal={.Ques.}}
pronoun 1: verb+什麼/何/啥 Case 3: if 什麼 use as an interrogative pronoun then 什麼/何/啥 =def: {event: participant={.Ques.}}	吃什麼 <i>chi she me</i> ‘eat what’ def: {eat 吃: patient={.Ques.}} 說啥 <i>shuo sha</i> ‘talk what’ def: {speak 說: content={.Ques.}}
pronoun 2: verb+什麼/何/啥 or 什麼/啥+verb Case 4: if 什麼 use as an indefinite pronoun Then 什麼/何/啥={event: participant={x}} or {event: participant={entity: quantity={all}}}	拿什麼都可 <i>na she me dou ke</i> ‘It’s OK to get anything’ def: {hold 拿: patient={entity: quantity={all}}} 什麼也不怕 <i>she me ye bu pa</i> ‘Be afraid of nothing’ def: {.not.fear 害怕: cause={entity: quantity={all}}}

怎麼 *ze me*

怎麼 *ze me* ‘how/how come’ plays the role of adverb. It asks the value of an adverbial type of semantic role including means/method (How) and reason (Why). Two meanings ‘how’ and ‘why’ of 怎麼 *ze me* can roughly be discriminated by the telicity of matrix verbs in context. That is, 怎麼 *ze me* in a sentence with telic verb or verb phrase refers to event that have endpoints means ‘why’. Contrarily, 怎麼 *ze me* in an atelic sentence means ‘how’. The disambiguation rules are as below:

- (30) if 怎麼 *ze me* +event[-telic] then 怎麼=def: means={ .Ques. }
if 怎麼 *ze me* +event[+telic] then 怎麼=def: reason={ .Ques. }

Take following sentences as examples:

- (31) 怎麼吃 *ze me chi* ‘how to eat’
def: {eat|吃:means={ .Ques. }}
該怎麼做 *gai ze me zuo* ‘how to do’
def: {do|做:means={ .Ques. }}
要怎麼存活 *yao ze me cun huo* ‘how to survive’
def: {survive|存活:means={ .Ques. }}
- (32) 怎麼吃了 *ze me chi le* ‘why did you eat it’
def: {eat|吃:reason={ .Ques. }}
怎還不走 *ze hai bu zou* ‘why haven’t you left’
def: { .not.leave|離開:reason={ .Ques. }}
怎麼沒去 *ze me mei qu* ‘why haven’t you gone’
def: { .not.go|去:reason={ .Ques. }}
怎麼會走 *ze me hui zou* ‘why did you go’
def: {leave|離開:reason={ .Ques. }}

In Mandarin Chinese, 了 *le*, 還 *hai* and the negative words 沒 *mei*, 不 *bu* all mark the endpoint of event; therefore, we use them as constraints to disambiguate the senses of 怎麼 *ze me*. Additionally, when a modal word appears between 怎麼 *ze me* and the event, normally the event also has an endpoint, that is, 怎麼 *ze me* questions ‘reason’ here. Unlike

this, when 怎麼 *ze me* appears between the modal word and event, the event does not have an endpoint, so 怎麼 *ze me* questions ‘means’ instead. Their differences can be found in example (33).

- (33) 他怎麼可以去 *ta ze me ke yi qu* ‘why can he go’
 def: {go|去:agent={he|他},reason={.Ques.}}
 他可以怎麼去 *ta ke yi ze me qu* ‘how can he go’
 def: {go|去:agent={he|他},means={.Ques.}}

怎樣 *ze yang* /如何 *ru he*

怎樣 *ze yang* plays the role of adverb. Its synonym is 如何 *ru he*. There are four senses of these words, *i.e.* ‘how’, ‘what’, ‘what is’, and ‘how about’. Their sense disambiguation rules are shown in (34):

- (34) if 怎樣 *ze yang* /如何 *ru he* + event[-telic] then def:means={.Ques.}
 if verb+怎樣 *ze yang* /如何 *ru he* then def:participant={.Ques.}
 if semantic role+怎樣 *ze yang* /如何 *ru he* then def:role={.Ques.}
 if entity+怎樣 *ze yang* /如何 *ru he* then def:truth={.Ques.}

Examples are shown in (35).

- (35) 怎樣採買 *ze yang cai mai* ‘how to buy’ def: {buy|買:means={.Ques.}}
 如何游泳 *ru he you yong* ‘how to swim’ def: {swim|游:means={.Ques.}}
 打算怎樣 *da suan ze yang* ‘intend to what’ def: {plan|計畫:content={.Ques.}}
 變成怎樣 *bian cheng ze yang* ‘become what’ def: {become|成爲:result={.Ques.}}
 味道怎樣 *wei dao ze yang* ‘what’s the taste’ def: taste={.Ques.}
 結果如何 *jie guo ru he* ‘what’s the result’ def: result={.Ques.}
 明天怎樣 *ming tian ze yang* ‘how about tomorrow’ def: truth={.Ques.}

In the case of entity+怎樣 *ze yang* /如何 *ru he*, it usually happens when the main verb of the sentence is omitted. In such cases, we must first recover the omitted part of the surface sentence based on the context then infer the complete sense. For example,

(36) 明天怎樣 *ming tian ze yang* ‘how about tomorrow’

Restore ellipsis:

明天一起去怎樣 *ming tian yi qi qu ze yang* ‘how about go together tomorrow’

def: {go|去:manner={together|共同},time={tomorrow|明天},willingness={.Ques.}}

However, the method for recovering the omitted part of the surface sentence is out of the scope of this paper.

多 *duo*

多 *duo* ‘how’ also plays the role of an adverb. It’s usually followed by an attribute value, such as 甜 *tian* ‘sweet’, 聰明 *cong ming* ‘smart’, 遠 *yuan* ‘far’, 大 *da* ‘big’. It can be used to express feelings of exclamation or doubt. We can not simply distinguish these two senses by the context, but we need to rely on the tone. For this reason, we will deal only with the senses of doubt. Incidentally, it is always possible to turn a declarative statement into a question by using a slightly rising intonation pattern. For the same reason, we do not deal with such sentences and a few interrogative words, such as 啊 *a*, 吧 *ba*, 呢 *ne*, as well.

多 *duo* ‘how’ with interrogative sense can be represent as below:

(37) 多 *duo* +attribute value def: {attribute value: degree={.Ques.}}

The real examples are:

(38) 多甜 *duo tian* ‘how sweet’

def: {sweet|甜:degree={.Ques.}} = def: sweetness={.Ques.}

多聰明 *duo cong ming* ‘how smart’

def: {smart|聰明:degree={.Ques.}} = def: smartness={.Ques.}

多遠 *duo yuan* ‘how far’

def: {far|遠:degree={.Ques.}} = def: distance={.Ques.}

多大 *duo da* ‘how big’

def: {big|大:degree={.Ques.}} = def: size={.Ques.}

The semantic roles of ‘sweetness’, ‘smartness’, ‘distance’, and ‘size’ are inferred from their respective values ‘sweet|甜’, ‘smart|聰明’, ‘far|遠’, and ‘big|大’ by checking the taxonomy of feature-value hierarchy in E-HowNet.

哪 *na*

幾 *ji*

For the disambiguation of 哪 and 幾, please see Section 3.3 (19) rules (E) and (F).

5. Conclusion and Future Works

To achieve near canonical semantic representation, we studied the semantic representation and composition of interrogatives. According to the semantic classification of interrogatives, we represent interrogatives in a hierarchy as follows:

true/false interrogative	def: truth={ .Ques. }
Wh-word interrogative	
asking factual information	def: role={ .Ques. }
asking relationship	def: . Ques.RelationRole()
asking opinion	def: ModalityRole={ .Ques. }
choosing options	def: role={ .option. { {x}.or. {y} } }

We have cited two examples (2), (3) earlier to illustrate automatic natural language understanding. After the discussion above, we can examine the result of this work:

(2) 我能否拍照？

Wo neng fou pai zhao?

Is it OK for me to take pictures?

Representation:

我 <i>wo</i> 'I'	def: {speaker 說話者}
能否 <i>neng fou</i> 'is it ok'	def: allowance={ .Ques. }
拍照 <i>pai zhao</i> 'take picture'	def: {TakePicture 拍攝}

Composition:

def: {TakePicture|拍攝:agent={speaker|說話者}, allowance={ .Ques. } }

(3) 我可不可以照相？

Wo ke bu ke yi zhao xiang?

Can I take photos?

Representation:

我 *wo* ‘I’ def: {speaker|說話者}

可不可以 *ke bu ke yi* ‘can’ def: allowance={.Ques.}

照相 *zhao xiang* ‘take photo’ def: {TakePicture|拍攝}

Composition:

def: {TakePicture|拍攝:agent={speaker|說話者}, allowance={.Ques.}}

Although the surface structures of (2), (3) are different, we find that their result of composition is the same. It means that, by means of E-HowNet representation, a machine can judge the sense similarity of words, phrases, and sentences and achieve machine understanding. However, this is only an illustration by example. For future research, we will implement a parsing system incorporated with the E-HowNet model to perform semantic composition process practically. To achieve this goal, apart from sense disambiguation, we find that discordance between syntactic structure and semantic relations is another critical problem. Take Sentence (39) as an example:

(39) 長途旅行不是很辛苦嗎？

Chang tu lu xing bu shi hen xin ku ma?

Isn't it hard for Long-distance travel?

Its parsing result is:

Theme[VP:(manner[A:長途 *long distance*]+Head[VA4:旅行 *travel*])+negation[Dc:不 *not*]+epistemics[Dbaa:是 *be*]+degree[Dfa:很 *very*]+Head[VH16:辛苦 *hard*]+particle[Td:嗎 *ma*]

The E-HowNet sense representation of (39) is:

def: {hard|辛苦:theme={travel|旅行:distance={far|遠}},degree={very|很},truth={.Ques.}}

Comparing the semantic representation with syntactic structure, we find rhetorical interrogative 不是嗎 *bu shi ma* ‘Isn't it’ is segmented into three words in syntax analysis, but

in the semantic point of view, they are integrated into one word and represented as ‘truth={.Ques.}’. There are still many types of discordance between syntactic structure and semantic relations that need to be studied. Furthermore, we have to find the mapping rules and match coarse-grained syntactic arguments to fine-grained semantic relations in the future. These results are applicable to both declarative sentences and interrogative sentences.

In conclusion, this study sheds new light on designing better and accurate question-answering systems because E-HowNet representation of questions not only represents their senses, but also marks the focused information to be answered. In addition, the proposed representational scheme also provides a way to convert a given sentence into a near-canonical sense representation. Therefore, the design of Chinese QA system will be our future task.

Acknowledgement

This research was supported in part by the National Science Council under a Center Excellence Grant NSC 96-2752-E-001-001-PAE and Grant NSC96-2221-E-001-009.

References

- Chen, K. J., S. L. Huang, Y. Y. Shih, and Y. J. Chen, “Extended-HowNet: multi-level concept definition and complex relation description,” 2004, Peking University.
- Chen, K. J., S. L. Huang, Y.Y. Shih, and Y. J. Chen, “Extended-HowNet- A Representational Framework for Concepts,” In *Proceedings of IJCNLP-05 Workshop on Lexical Semantic*, 2005, Jeju Island, South Korea.
- Chen, K. J., et al., Sinica Treebank, version 3.0, <http://turing.iis.sinica.edu.tw/treesearch/>, 2005.
- Chung, Y. S., S. L. Huang, and K. J. Chen, “Modality and Modal Sense Representation in E-HowNet,” In the *Proceeding of the 21st Pacific Asia Conference on Language, Information and Computation (PACLIC)*, 2007, Seoul National University, pp. 136-145.
- Dong, Z. D., HowNet, <http://www.keenage.com/>, 1988.
- Ginzburg, J., and I. A. Sag, *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*, Stanford: CSLI Publications, 2000.
- Li, C. N., and S. A. Thompson, *Mandarin Chinese*, The Crane Publishing Co., LTD., 1997.
- 石定栩(Shi, Dingxu), *共性與個性—漢語語言學中的爭議*, 北京語言文化大學出版社, 北京, 1999.
- 呂叔湘(Lv, Shuxiang), *中國文法要略*, 文史哲出版社, 台北, 1942.
- 邵敬敏(Shao, Jingmin), *現代漢語疑問句研究*, 華東師範大學出版社, 上海, 1996.
- 馬建忠(Ma, Jianzhong), *馬氏文通*, 商務印書館, 台北, 1935.

- 陸儉明(Lu, Jianming), “關於現代漢語裡的疑問語氣詞”, 收錄於 *中國語文*, 第五期, 1984, pp. 330-332.
- 湯廷池(Tang, Tingchi), “國語疑問句的研究”, 收錄於 *師大學報*, 第二十六期, 1981, pp. 1-59.
- 黎錦熙(Li, Chihsi), *新著國語文法*(Chinese grammar on the national language), 商務印書館, 台北, 1930.

