Lexical Semantic Representation and Semantic Composition

An Introduction to E-HowNet

Content

Preface

1. Introduction
   1.1 Lexical knowledge representation—WordNet and HowNet’s approach
      1.1.1 WordNet approach
      1.1.2 HowNet approach

2. E-HowNet
   2.1 Taxonomy & Ontology
      2.1.1 Primitive—Entities and Relations
      2.1.2 The usages of E-HowNet’s Relations
      2.1.3 The difference between E-HowNet Ontology and HowNet Ontology
   2.2 Principles for sense definition
      2.2.1 A concept is defined by its hypernym and prominent properties
      2.2.2 Use well-defined/primitive concepts and relations to define new concepts
      2.2.3 Multi-level representations: High-level representations can be decomposed into primitive representations
   2.3 Representations for different types of senses
      2.3.1 Content senses
      2.3.2 Relational senses
   2.4 Syntax of E-HowNet expressions

3. Advantages of E-HowNet

4. Semantic Composition and Decomposition

5. Potential Applications of E-HowNet
   5.1 Identify senses of new compound words
   5.2 Sense disambiguation
   5.3 Semantic role assignment
   5.4 Filling semantic gaps by automatic deduction
   5.5 Toward near-canonical meaning representation

6. Conclusions and Future Research

References

Appendix A. Formal Syntax of E-HowNet Expressions
Preface

The purpose of designing the lexical semantic representation model E-HowNet is for natural language understanding. Extended-HowNet (short as E-HowNet) is a frame-based entity-relation model extended from HowNet (Dong & Dong 2006) to define lexical senses (concepts). The following features are major extensions:

a. Word senses (concepts) are defined by not only primitives but also any well-defined concepts and conceptual relations.
b. A uniform sense representation for content words, function words and phrases
c. Semantic relations are explicitly expressed
d. Semantic composition and decomposition capabilities
e. Near-canonical representations for lexical senses and phrasal senses

The above features were set to serve the purpose of natural language understanding. We do not claim that we had achieved the goal already. Although the current version has achieved only coarse-grained representation, we believe that it has enough lexical coverage and is practically useful. We hope that the ultimate goal of natural language understanding will be accomplished after future improvement and evolution of the current E-HowNet.

The development of E-HowNet started in 2003. We would like to thank Dr. Dong who had laid the foundation of this lexical sense representation model, i.e. HowNet, and generously allowed us to build E-HowNet based on his original establishments. Most of the lexical sense representations of E-HowNet were revised or adopted from HowNet. The set of primitives (called sememes in HowNet) and their taxonomy were also retained and adjusted to suit the goal of semantic composition. Due to its open-ended nature, it is always possible for a conceptual representation to be refined by replacing coarse-grained knowledge with fine-grained knowledge. We will continue to improve our representations and correct possible errors in the future.

We would like to thank Shu-Ling Huang, Yueh-Yin Shih, Yi-Jun Chen, Su-Chu Lin, You-Shan Chung, Ming-Hong Bai who contributed to the development and design of E-HowNet.

Keh-Jiann Chen

CKIP group http://rocling.iis.sinica.edu.tw/CKIP/

Institute of Information Science, Academia Sinica

September 2009
Introduction

E-HowNet is an entity-relation model that represents lexical senses. It was extended and evolved from HowNet (Dong & Dong 2006). HowNet is an on-line common-sense knowledge-base indexing relations of concepts obtained from lexicons of Chinese and English. Each concept is represented and understood by its definition and associated links with other concepts. HowNet’s lexical sense definitions provide more information than WordNet’s hyponymy relations. They also encode relational links between words via feature relations. HowNet has the following advantages over WordNet: a) inherent properties of concepts are derived from encoded feature relations in addition to hypernymous concepts, and b) information regarding conceptual differences between different concepts and information regarding morpho-semantic structure are encoded. HowNet’s advantages make it an effective electronic dictionary for NLP applications. In recent years, HowNet has been applied to a variety of research topics including: (1) word similarity (劉 & 李 2002), (2) machine translation and (3) information retrieval etc.

However, what interests us here is how to use HowNet to achieve mechanical natural language understanding. When we say that a sentence is ‘understood’, we mean that the concepts and the conceptual relationships expressed by the sentence are unambiguously identified, and we can make correct inferences and/or responses. Therefore to achieve natural language understanding, computer systems should know the sense similarity and dissimilarity of words and sentences. A representational framework which represents knowledge about lexical concepts and performs the following functions is needed.

a) Identifies synonymous concepts and measures similarity distance between two concepts (劉 & 李 2002).

b) Knows the shared semantic features and feature differences between two concepts.

c) Provides unique indices to each concept, such that associated knowledge can be coded and accessed.

d) Language independent sense encoding.

e) Logical inferences through conceptual property inheritance system.

f) Dynamic concept decomposition and composition mechanisms.

None of the currently available ontology provides all of the above functions and so far there has been little research on applying HowNet to semantic composition. We therefore extend HowNet to deal with this problem. The resulting system is called E-HowNet.

1.1 Lexical knowledge representation—WordNet and HowNet’s approach

Words are the smallest meaningful units of a language which serve as indices to access various knowledge,
such as grammatical functions, semantic knowledge and world knowledge. On account of sense ambiguity, one word may have more than one sense, with each associated with a set of syntactic, semantic, and world knowledge information. The form shown as (1)

\[
\begin{align*}
\text{Word} : & \quad \text{sense1 : grammatical function } \quad \text{semantic knowledge} \quad \text{world knowledge} \\
& \quad \text{sense2: grammatical function } \quad \text{semantic knowledge} \quad \text{world knowledge} \\
& \quad \text{sense3: …}
\end{align*}
\]

1.1.1 WordNet approach

WordNet (Fellbaum, 1998) contains information about nouns, verbs, adjectives and adverbs in English and is organized around the notion of a synset. A synset, roughly denoting a concept, is a set of words with the same part-of-speech that can be interchanged in a certain context. For example, \{car; auto; automobile; machine; motorcar\} form a synset because they can be used to refer to the same concept. Synsets can be related to each other by semantic relations, such as hyponymy, meronymy, cause, etc. and a synset is often further described by a gloss: “4-wheeled; usually propelled by an internal combustion engine.”

Synsets can be related to each other by semantic relations. Table 1 contains some examples:

<table>
<thead>
<tr>
<th>Synset</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>vehicle</td>
<td>{conveyance; transport}</td>
</tr>
<tr>
<td></td>
<td>{vehicle}</td>
</tr>
<tr>
<td>motor</td>
<td>{motor vehicle; automotive vehicle}</td>
</tr>
<tr>
<td></td>
<td>{motor vehicle; automated vehicle}</td>
</tr>
<tr>
<td>car</td>
<td>{car; auto; automobile; machine; motorcar}</td>
</tr>
<tr>
<td></td>
<td>{cruiser; squad car; patrol car; police car; prowl car}</td>
</tr>
<tr>
<td></td>
<td>{cab; taxi; hack; taxicab;}</td>
</tr>
<tr>
<td></td>
<td>{hinge; flexible joint}</td>
</tr>
<tr>
<td></td>
<td>{bumper}</td>
</tr>
<tr>
<td></td>
<td>{door; doorlock}</td>
</tr>
<tr>
<td></td>
<td>{car door}</td>
</tr>
<tr>
<td></td>
<td>{car window}</td>
</tr>
<tr>
<td></td>
<td>{car mirror}</td>
</tr>
<tr>
<td></td>
<td>{arm rest}</td>
</tr>
</tbody>
</table>

The disadvantage of WordNet-like ontology is that each concept class has limited linking to other concepts. The major links are hyponymy relations which limit inheritance and inference capability to the classes on the taxonomy. For those features not used as classification criterion, it is not possible to encode their inherent properties. For instance, the set of round objects, edible things will not be a natural class in the taxonomy. Therefore there will not be any general inference rules, such as (roll, <round object>), (digest, <edible things>) encoded. In sum, WordNet’s approach does not provide information regarding conceptual differences between different synsets, information for unknown words, or mechanisms for semantic composition.
1.1.2 HowNet approach

HowNet is an on-line common-sense knowledge base unveiling the inter-conceptual relations and inter-attribute relations of concepts conveyed by Chinese words and their English equivalents (Dong & Dong 2006). Compared with 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) inherent properties of concepts are derived from encoded feature relations in addition to hypernymous concepts, and (b) information regarding conceptual differences between different concepts and information regarding morpho-semantic structure are encoded. HowNet’s advantages make it an effective electronic dictionary for NLP applications.

Conventional sense representation have used semantic primitives to define and achieve canonical representation for concepts (Wierzbicka, 1972), such as Conceptual Dependency representation (Schank, 1975) and HowNet. However, using primitives to define concepts causes information degrading as it is almost impossible to understand a definition of a complex concept merely with primitives. Furthermore, it is debatable whether there exists a limited and fixed set of so-called primitives. In HowNet, word sense definition is restricted to a set of around two thousands primitive concepts, called sememes. A word sense is defined by its hypernymous sememe and additional semantic features. For instance, the HowNet definition of Warrior|戰士 is as (1):

(1) \{human|人:belong={army|軍隊},
\{fight|爭鬥:
\agent={~},
\domain={military|軍}\}\}

The representation says that a warrior is a human in the army who plays the role of agent in the event of military fighting.

HowNet describes the following conceptual relations:

- **Hypernymy** 上下位關係
- **Synonymy** 同義關係
- **Antonymy** 反義關係
- **Attribute-host** 屬性-宿主關係
- **Part-whole** 部件-整體關係
- **Event-role** 事件-角色關係
HowNet Ontology is as (3):

(3)     V  event|事件
        V1  static|靜態     V2  act|行動
        V1.0  relation|關係     V2.0  AlterRelation|變關係
        V1.01  isa|是非關係    V2.01  AlterIsa|變是非
        ...
        V1.1  state|狀態     V2.1  AlterState|變狀態

Common sense knowledge is also partially encoded in HowNet and is exemplified in the conceptual graph of Figure 1 quoted from HowNet (http://www.keenage.com/).

Figure 1. An example of the lexical representations of HowNet

The disadvantages of the HowNet approach are:

- Representation by primitives degrades precision and readability.
  
  i. 老虎 tiger def:{beast|走獸} and 熊 bear def:{beast|走獸};
  
  ii. 鉗子 forceps def:{tool|用具}:{hold|拿:instrument={~}};
  
  iii. 鐘錶店 watchmaker's shop
Semantic relations are not explicitly expressed.

Sense of function words and relational concepts are not well established.

e.g. function words: 僅 just

Semantic composition and decomposition are not taken into consideration.
2. E-HowNet

The purpose of the lexical semantic representation model E-HowNet is for natural language understanding. E-HowNet is a frame-based entity-relation model extended from HowNet (Dong & Dong 2006) to define lexical senses (concepts), and it intends to achieve the following goals (Chen et al., 2004; Chen K.J., Huang, Shih & Chen Y.J., 2005; Chen Y.J., Huang, Shih & Chen K.J., 2005; Huang, Chung & Chen, 2008):

- Word senses (concepts) are defined by not only primitives but also any well-defined concepts and conceptual relations.
- A uniform representation model for function words and content words, as well as phrases.
- Semantic relations are explicitly expressed for all meaning representations.
- Semantic composition and decomposition capabilities.
- Near canonical representations for lexical senses and phrasal senses.

The E-HowNet system comprises the following components:


b. A set of primitive concepts, called sememes in HowNet which include events, objects, relations, and values (Also see http://ckip.iis.sinica.edu.tw/taxonomy/)

c. The E-HowNet expressions for all lexical senses of CKIP word entries

2.1 Taxonomy & Ontology

To achieve natural language understanding, computer systems should know the sense similarity of and the dissimilarity between two sentences or two words. To achieve the above goals, it requires the support of ontologies. Ontology provides the following functions.

- Identifies synonymous concepts and measures similarity distance between two concepts (劉 & 李, 2002).
- Knows the shared semantic features and feature differences between two concepts.
- Provides unique index to each concept, such that associated knowledge can be coded and accessed.
- Language independent sense encoding
- Logical inferences through conceptual property inheritance system
- Dynamic concept decomposition and composition mechanisms

In E-HowNet all concepts are either primitive concepts or defined by simpler concepts in terms of an entity-relation model (Chen et al., 2004; Chen K.J., Huang, Shih & Chen Y.J., 2005; Chen Y.J., Huang, Shih &
Chen K.J., 2005; Huang, Chung & Chen, 2008). In the examples of this paper, a primitive concept will have an English equivalent beside it, e.g. {read|讀}, whereas simpler concepts will be defined by Chinese words only, e.g. {狗}.

The concepts form a hierarchical structure by is-a (hyponymy) relations, as shown in [http://ckip.iis.sinica.edu.tw/taxonomy/](http://ckip.iis.sinica.edu.tw/taxonomy/). It is obvious that the associated property or knowledge regarding a particular concept can be directly accessed or encoded through its definition or indirectly inherited from its ancestors. Furthermore, the hierarchical taxonomy also indicates the semantic distance between two concepts. However, conventional taxonomies do not provide the exact semantic similarities and dissimilarities of two concepts. In E-HowNet, definitions of concepts show not only the semantic similarities of two concepts but also the semantic differences between them. For instances, <teacher> and <student> are both <human> and hence inherit the properties of <human>. They also participate in the event of <teach>, but the semantic difference is that they are denoted by different semantic roles and therefore inherit different properties of their semantic relations.

Taxonomically unrelated but conceptually related concepts can also be computably associated through their E-HowNet definitions. Figure 1 in section 1 shows that the concepts are not only linked by taxonomical relations but also linked by other semantic relations.

2.1.1 Primitives—Entities and Relations

There are about two thousand and six hundred primitives, forming a taxonomy comprised of two types of subtrees of entities and relations. The entities include events, objects, and attribute-values. The relations include semantic-roles and functions. Entities indicate concepts that have substantial content. By contrast, relations play the role of linking semantic relations between entities (Chen et al., 2004; Chen K.J., Huang, Shih & Chen Y.J., 2005; Chen Y.J., Huang, Shih & Chen K.J., 2005; Huang, Chung & Chen, 2008).

There are two different types of relations, semantic roles and functions. All semantic roles are binary relations rel(x,y), with the parameter x usually being the dependency head. We write rel(x,y) as rel(x)={y}, which reads as ‘rel of x is y’. For example, agent(eat)={dog} means ‘agent of eating is a dog’. The sense of the event ‘Dog eats’ is expressed as {eat: agent={dog}} in E-HowNet, where ‘agent={dog}’ is an abbreviation of agent(～)={dog} and ～ denotes the head concept, which is ‘eat’ in this example. A relation rel(x)={y} is considered as a mapping from domain(x) to range(y). The value of domain and range depends on the relation type. In HowNet the ranges of attribute types of relations are their attribute-values. For instance, the color-values are blue|藍, red|紅, green|綠 and so forth. Another kind of semantic roles is participants of events, such as agent, theme, goal etc. Their range values are determined by the head events.

Function is a special kind of relation in which a concept is mapped onto another concept of the same domain. Rather than establishing the thematic relation or property attribute between two parameters, functions

---

transform a concept to a new concept. Function has compositional property. New functions can be constructed by combining functions of the same domain. For instances, the kinship function of Father(Father(x)) denotes ‘the grandfather of x’ and the direction function of north(east()) denotes ‘the direction of north-east.’ Both are compositions of basic functions. Function expressions are written as rel(x) and treated as a concept or sememe in E-HowNet expression. (4)-(6) are typical examples.

(4) vehicle headlight 車燈
   def: {PartOf({LandVehicle|車}): telic={illuminate|照射}: instrument={~}}).
(5) father-in-law 岳父/公公
   def: {father({spouse({human|人})})}.
(6) Eastern Taiwan 東台灣
   def:{east({Taiwan|台灣})}

In (4), ‘PartOf’ is a function while ‘telic’ and ‘instrument’ are semantic roles. ‘Telic’ relates the target object to the event, so does ‘instrument.’ By contrast, ‘part of’ does not relate entities of different domains but marks the range of the target object.

In E-HowNet, we also regard and-or relation, question and negation relation as functions (Huang & Chen 2008; Chen Y.J., Huang, Shih & Chen K.J., 2005). Their usage is shown as follows:

(7) get in and out 進出
   def: {or({GoInto|進入},{GoOut|出去})}.
(8) why 為何
   def: question({cause({event|事件})}).
(9) be distracted 分神
   def: {not({attentive|專注})}.

However, semantic roles also have the form rel(x) which signals an underspecified value to be filled to complete the expression. Below are some examples:

(10) a. wavelength 波長
    def: length({phenomena|現象:cause={shiver|顫動}})

b. wavelength 10 km 波長十公里
    def: length({phenomena|現象:cause={shiver|顫動}})={10 公里}
c. electric wave which has a wavelength of 10 km 波長十公里的電波

def: \{length\{phenomena\{現象\:cause\{shiver\:顫動\:agent\{electricity\:電\}\}\}\}\}=\{10 \text{公里}\}\}

In order to achieve automatic feature unification processes, we organized relations into a hierarchical structure similar to the taxonomy for entities. A hyponym relation entails its hypernym relations. Their usages are demonstrated in the next section.

2.1.2 The usages of E-HowNet’s Relations

A. Semantic roles for objects

quantifier—expresses a definite or indefinite amount of quantity, e.g. 七星山 def:
{山\:quantifier\{definite\:定指\},name\{"七星山"\},location\{Taiwan\:台灣\}}

quantity—the quantity of an object, e.g. 人群 def: \{human\:人\:quantity\{many\:多\}\}

container—the container of an object; defines measure words (Tai et al., 2009), e.g. 篮
def:container={篮子}

ordinal—the ordinal of object, e.g. 甲 def:ordinal={1}

possessor—the owner in a possessor-possession relation, e.g. 失主 def: \{human\:人\:predication\{lose\:失去\:possessor\{~\}\}\}

member—the member of an object, e.g. 女校 def: \{學校\:member\{女生\}\}

creator—the creator of an object, e.g. 自序 def: \{text\:語文\:telic\{describe\:描寫\:manner\{concise\:扼要\},instrument\{~\}\},creator\{作者\}\}

owner—the owner of an object, e.g. 官房 def: \{宿舍\:owner\{政府\}\}

whole—the object which is the whole of its parts, e.g. 人頭 def: \{head\:頭\:whole\{human\:人\}\}

predication—the event in which the head object participates, e.g. 大鬍子 def: \{human\:人\:predication\{own\:有\:possession\{鬍子\:quantity\{多\}\},agent\{~\}\}\}

telic—purpose and function of an object, e.g. 甘蔗田 def: \{田\:telic\{planting\:栽植\:patient\{ crop\:莊稼\:taste\{sweet\:甜\}\},location\{~\}\}\}

agentive—factors involved in the origin or “bringing about” of an object, e.g. 丹藥 def: \{medicine\:藥物\:agentive\{提煉\:PatientProduct\{~\},agent\{~\},predication\{teach\:教\:agent\{~\},content\{natural\:天然\}\}\}\}

scene—the scene of event

locational—the locality of an entity
**location**—the location where an event happens, e.g. 水鳥 def: {bird|禽:location={waters|水域}}

**LocationIni**—the starting point of a trajectory over which an event takes place

e.g. 起跑點 def: LocationIni({run|跑})

**LocationFin**—the endpoint of a trajectory over which an event takes place

e.g. 終點 def: {place|地方:predication={arrive|到達}:LocationFin={~}}

**LocationThru**—the trace of a trajectory over which an event takes place

e.g. 航行路徑 def: LocationThru({航行})

**direction**—the direction to which an object faces, e.g. 台灣以東 def: direction={east({Taiwan|台灣})}

**distance**—the distance between an object and a reference point, e.g. 九霄 def: {sky|空域:distance={far|遠}}

**position**—the position where an object occupies, e.g. 東台灣 def: {PartOf({Taiwan|台灣}):position={east({Taiwan|台灣})}}

**source**—the source of an entity, e.g. 內憂 def: {experience|感受:CoEvent={unfortunate|不幸}:source={internal({country|國家})}}

**temporal**—the relations concerning times

**duration**—a period of time for which a situation lasts, e.g. 分秒 def: {time|時間:duration={TimeShort|短時間}}

**while**—an event that occurs at the same time as the main event,

**TimePoint**—a specific period of time in the past, present or future where an object belongs to,

e.g. 古蹟 def: {building|建築物:TimePoint={past|過去}}

**TimeFeature**—the status of a time point in relation to a reference time, e.g. 月底 def: {month|月:TimeFeature={ending|末}}

**domain**—the domain of an entity, e.g. 砂盤 def: {tool|用具:telic={plan|計}:instrument={~}:domain={military|軍}}

**property**—the property of an object.

**constitutive**—the relation between an object and its constituents, such as material, components etc.
content— the content of an object, e.g. 二十四孝 def: {故事:content= {emotion|情感:CoEvent={loyal|忠孝}} }

material— the material of an object, e.g. 土牆 def: {牆:material={stone|土石}}

component— the component of an object, e.g. 石灰礦   def: {礦物:component= {material|材料:telic={build|建造:material={~}}} }

ingredients— the ingredients of an object, e.g. 綠豆湯 def: {湯:ingredients={綠豆}}

formal— what distinguishes the object within a larger domain, such as shape, magnitude, color etc..

CoEvent— introduces the event type of an event noun, e.g. 大戰 def: {fact|事情:CoEvent={fight|爭鬥}}

name— the name of an object; the symbol ‘”’ indicates that it is a character string, e.g. 巴哈 def: {作曲家:quantifier={definite|定指},name= {“巴哈”},location={Germany|德國}}

gender— the gender of a living thing, e.g. 女伶 def: {藝人:gender= {female|女}}

evaluation— evaluations of objects

HumanPropensity— personalities of a person, e.g. 正派人 def: {human|人:HumanPropensity={decent|正經}}

qualification— non-measureable properties of an object, e.g. 人造花 def: {FlowerGrass|花草:qualification={artificial|人為}}

attribute— the physical properties of an object, e.g. 乾飯 def: {飯:attribute= {StateSolid|固態}}

age— the age of a living thing, e.g. 幼童 def: {human|人:age={child|少兒}}

color— the color of an object, e.g. 丹楓 def: {葉:color={ red|紅}}

speed — moving speed of an object, e.g. 疾行 def: {walk|走:speed={fast|快}}

weight— the weight of an object, e.g. 公克 def: weight={公克}

length— the length of an object, e.g. 樁子 def: {tool|用具:length={LengthLong|長},shape={straight|直}}

kind — object types, e.g. 鳥類 def: kind={bird|禽}

volume— volume of an object, e.g. 公石 def: volume={公石}

2 Due to limited space, among the semantic roles HumanPropensity, qualification and attribute, only some of the hyponyms of the last one are listed here. For the complete taxonomy, please refer to http://ckip.iis.sinica.edu.tw/taxonomy/.
height—height of an object, e.g. 高廊 def: {passage(建築物):height={high|高}}
width—width of an object, e.g. 窄巷 def: {巷子: width={narrow|窄}}
size—size of an object, e.g. 血盆大口 def: {mouth: size={big|大}}
taste—taste of an object, e.g. 苦酒 def: {酒:taste={bitter|苦}}
temperature—temperature of an object, e.g. 熟食 def: {food:temperature={hot|熱}}
form—object forms,
  appearance—appearances of an object, e.g. 人形 def: appearance={human|人}
  dimension—dimensionality of an object, e.g. 木塊 def: {wood:dimension={cubic|體}}
  shape—shape of an object, e.g. 球 def: {object:shape={round|圓}}

B. Semantic roles for events

scene—scene of events

situation—event situations

  instrument—an object which is used as a tool in an event, e.g. 娛樂片 def: {影
  片:telic={recreation|娛樂};instrument={~}}}

frequency—the frequency an event happens, e.g. 三顧茅廬 def: {invite;邀請;frequency={again|再}}}

degree—the scale of intensity or quality, e.g. 一塵不染 def: {spotless:潔;degree={extreme|極}}

manner—the way an event happens, e.g. 冷笑 def: {laugh;笑;manner={wicked|歹}}

means—the method how an act is done, e.g. 餵食 def: {eat;吃;means={feed|餵}}

condition—the condition under which an event happens e.g. 不藥而癒 def: {BeRecovered|復
  原;StateFin={sturdy|健壯},StateIni={ill|病態},condition={not({eat|吃});patient={medicine|藥物}}}

hypothesis—what is assumed about an event, e.g. 如果 def: hypothesis={}

concession—despite/although, e.g. 老當益壯 def: {healthy|康健};concession={aged|老年}}

conversion—otherwise, e.g. 否則 def: conversion={}

restriction—restrictions of an event, e.g. 不僅 def: restriction={}
alternative—an alternative way, e.g. 要么 def: alternative={}

selection—suggested option(s), e.g. 不如 def: selection={}

rejection—excluded option(s), e.g. 與其 def: rejection={}

avoidance—what is avoided, e.g. 以免 def: avoidance={}

whatever—no matter how/who/what/etc. e.g. 不論 def: whatever={}

environment— circumstances of an event

temporal— the relations concerning event times (Huang & Chen, 2009)

TimePoint—a specific time when an event happens, e.g. 換發 def: {issue|分發:
TimePoint={TimeAfter({check|查}),means={replace|代替}}}

TimeIni—a specific time when an event begins, e.g. 有始以來 def: TimeIni={past|過去}

TimeFin—a specific time when an event ends, e.g. 熬到 def: {endure|忍耐:TimeFin={}}

duration—the period of time for which an event lasts,

TimeFeature—the status of a time point in relation to a reference time e.g. 期中 def: {time|
時間:domain={education|教育},predication={study|學
習:duration={~}},TimeFeature={middle|間}}}

while—an event that occurs at the same time as the main event, e.g. 相視而笑 def: {laugh|
笑:while={look|看:manner={EachOther|相互}}}

locational

location—the location at which an event takes place, e.g. 人行道 def: {route|道
路:telic={walk|走:theme={行人},location={~}}}

LocationIni—the starting point of a trajectory over which an event takes place, e.g. 逃家
def: {flee|逃跑:LocationIni={family|家庭}}

LocationFin—the endpoint of a trajectory over which an event takes place, e.g. 升空 def:
{rise|上升:LocationFin={sky|空域}}

LocationThru—the trace of a trajectory over which an event takes place, e.g. 越線 def:
{cross|越過:LocationThru={mark|標誌:dimension={linear|線}}}

distance—the distance over which an act takes as its scope, e.g. 放眼望去 def: {look|
看:distance={far|遠}}
direction—the direction a movement takes, e.g. 拉上來 def: {pull}
拉:direction={upper({object(物體)})}

direction—the direction a movement takes, e.g. 子弟兵 def: {army|軍隊}
隊:source={家鄉}

source—the source from which a possession is obtained, e.g. 子弟兵 def: {army|軍隊}

story—details of an event

disjunctive—indicates a contrast, e.g. 但是 def: disjunctive={}

range—how thoroughly something is done, e.g. 博覽群籍 def: {read|讀}:range={extensive|泛}

addition—an event beside the main one, e.g. 况且 def: addition={}

except—exceptions, e.g. 除外 def: except={}

topic—topic of an event, e.g. 關於 def: topic={entity|事物}

standard—standard of rules, e.g. 論功行賞 def: {reward|獎勵}:standard={merit()}

purpose—the purpose of an event, e.g. 下馬威 def: {show|表現}:manner={powerful|有威力},purpose={persuade|勸說}

cause—the cause of an event, e.g. 人老珠黃 def: {become|成爲}:StateFin={ugly|醜}

result—the result caused by an event, e.g. 人去樓空 def: {leave|離開}:result={inactive|冷清}

contradictory—to summarize, e.g. 也就是說 def: conclusion={}

price—the price one pays to purchase something, e.g. 佣金 def: {payment|酬金}:telic={ recommend|推薦}:price={~}

StateIni—the initial state before an event happens, e.g. 返璞歸真 def: {resume|恢復}:StateFin={pure|純}

StateFin—the new state into which something changes after an event happens, e.g. 本土化 def:
{AlterState|變狀態}:StateFin={native|本土}

ComparativeQuantity—the quantity in comparison, e.g. 我比他高一個頭 def: {tall|高}:theme={speaker|說話者},contrast={他},ComparativeQuantity={head|頭}:quantity={1}{}

ComparativeAttribute—the attribute in comparison, e.g. 身高上，我比他高一個頭 def: {tall|高}:theme={speaker|說話者},contrast={他},ComparativeAttribute={stature({human|人})}
participant— participant roles of an event

theme—the entity about which a stative situation concerns; an object that moves; or an object which is moved or changed its state, e.g. 一覽無遺 def: {lucid|清晰}:theme={look|看}}；流淚 def: {flow|流}:theme={BodyFluid|體液}:whole={animate|生物},source={eye|眼}}；卸貨 def: {TakeAway|搬動}:theme={artifact|人工物}}

product— products of an event

PatientProduct—an object which comes into physical being after an event, e.g. 下蛋 def: {reproduce|生殖}:PatientProduct={蛋}}

ContentProduct—an object that is produced by artistic activities, e.g. 打稿 def: {compile|編輯}:ContentProduct={text|語文}}

possession—an object which is owned in a possessor-possession relationship, e.g. 售貨 def: {sell|賣}:domain={economy|經濟},possession={artifact|人工物}}

experiencer—a living thing who feels an emotion, e.g. 好戰份子 def: {human|人}:predication={FondOf|喜歡}:target={fight|爭鬥},experiencer={~}}

agent—the actor of an event, e.g. 人力車 def: {LandVehicle|車}:predication={pull|拉}:agent={human|人},theme={~}}

goal—an object which is affected or perceived,

cost— the object which is perceived, e.g. 見天日 def: {perception|感知}:content={lights|光}}

patient—the object which is affected, e.g. 殭及池魚 def: {damage|損害}:patient={human|人},qualification={guiltless|無罪}}

target—the goal which is not really affected, e.g. 祭祖 def: {salute|致敬}:target={human|人},modifier={forefathers|祖先}}

source—the source from which a possession is obtained, e.g. 羊 def: {livestock|牲畜}:telic={take|取}:possession={hair|毛},source={~}}

benefactor—the object which an event benefits, e.g. 公費生 def: {學生}:predication={pay|付}:agent={政府},benefactor={~}}

partner—an object which is correspondent to the main body in the events of ‘connective,’ ‘AlterConnection,’ ‘TimeOrSpace,’ or ‘HaveContest’ types etc. e.g. 手下敗將 def: {human|人}:predication={defeat|戰勝}:patient={~}}

contrast—an entity that corresponds to another entity in some way e.g. 過猶不及 def: {equal|相等}:contrast={insufficient|欠},theme={over|}}
pragmatic—functions of a speech

SpeakerAttitude—the attitude/viewpoint of the speaker, e.g. 不得好死 def: {die| 死:manner={accidental|偶發},attitude(speaker|說話者)=ExpressAgainst|譴責}}

particle—a kind of function words without particular senses, e.g. 呢, 哇, 呀, 喲, 罷了 def: {null}

aspect—aspect of an event, e.g. 搏病 def: {ill|病態:aspect={Vgoingon|進展}}

ExtendedModality—speaker’s evaluation of an event (Chung, Huang & Chen, 2007)

epistemic—possibility of an event

possibility—e.g. 胜券在握 def: {win|獲勝:possibility={extreme|極}}

impossibility—e.g. 马生角 def: {fact|事情:CoEvent={happen|發生:impossibility={extreme|極}}

deontic—speaker’s permission

allowance—e.g. 責無旁贷 def: {bear|承擔:allowance={extreme|極}}

disallowance—e.g. 毋庸赘言 def: {speak|說:manner={redundant|多餘},disallowance={ish|稍}}

ability—capability of the agent

capacity e.g. 無敵 def: {defeat|戰勝:domain={military|軍},capacity={extreme|極}}

incapacity e.g. 愛莫能助 def: {help|幫助:incapacity={ish|稍}}

volition—volitionality of the agent

willingness e.g. 願聞其詳 def: {know|知道:willingness={ish|稍}}

unwillingness e.g. 宁为鸡首 def: {defeated|輸掉:unwillingness={ish|稍}}

expectation—degree of expectedness

expectedness e.g. 不出所料 def: expectedness={extreme|極}

unexpectedness e.g. 受寵若驚 def: {uneasy|不安:cause={undergo|經受:content={like|愛惜},unexpectedness={ish|稍}}}

truth—whether something is true, e.g. 是否 def: truth={Ques()}

C. Function

Kinship—kinship relations

parents e.g. 雙親 def: {parents({x:human|人})}
father e.g. 乃翁 def: {father(listener|聽者)}

mother e.g. 外婆 def: {mother(mother(x:human|人))}

offspring e.g. 後裔 def: {offspring(x:human|人):quantity={mass|眾}}

son e.g. 皇子 def: {son(皇帝)}

daughter e.g. 長女 def: {daughter(x:human|人):age={aged|老年}:degree={most|最}}

sibling e.g. 手足 def: {sibling(x:human|人)}

brother e.g. 媳子 def: {son(brother(x:human|人))}

YoungerBrother e.g. 弟弟 def: {YoungerBrother(x:human|人)}

ElderBrother e.g. 兄長 def: {ElderBrother(x:human|人)}

sister e.g. 從母 def: {sister(mother(x:human|人))}

YoungerSister e.g. 舍妹 def: {YoungerSister(speaker|說話者)}

ElderSister e.g. 女兄 def: {ElderSister(x:human|人)}

cousin e.g. 表哥 def: {cousin(x:human|人):age={aged|老年},gender={male|男}}

spouse e.g. 親家 def: {parents(spouse(offspring(x:human|人)))}

husband e.g. 良人 def: {husband(x:human|人)}

wife e.g. 弟妹 def: {wife(YoungerBrother(x:human|人))}

OtherHumanRelation

friend e.g. 令友 def: {friend(listener|聽者)}

self e.g. 自主 def: {decide|決定}:agent={self(x)}

member e.g. 兵 def: {member(army|軍隊)}

Direction/Position Value

east e.g. 台灣以東 def: direction={east(Taiwan|台灣)}

west e.g. 西北郊 def: {place|地方}:position={north(west(edge|邊):whole={city|市})}

north e.g. 川北 def: {place|地方}:position={north(provincial|省):name="四川",location={China|中國},quantifier={definite|定指}}

south e.g. 南部 def: {place|地方}:position={south(x:place|地方)}}
right e.g. 右側 def: direction={right({x})}

left e.g. 左左 def: {place|地方}:position={left({山|})}

upper e.g. 上呼吸道 def: {nerve|絡}:telic={respire|呼吸}:instrument={~}:position={upper({respire|呼吸}:instrument={~})}

beneath e.g. 地下 def: {place|地方}:position={beneath({land|陸地})}

WholePlace e.g. 郊區 def: {edge|邊}:WholePlace({city|市})

centre e.g. 掌心 def: {hand|手}:position={centre({hand|手})}

InBetween e.g. 頰 def: {place|地方}:position={InBetween({鼻|,}{耳|})}

side e.g. 一旁 def: {place|地方}:position={side({x: place|地方})}

InFront e.g. 臉 def: {head|頭}:position={InFront({head|頭})}

hind e.g. 舌後 def: {mouth|口}:position={hind({PartOf({mouth|口}):telic={or({發聲}:instrument={~}),{舐}:instrument={~}}})}

surrounding e.g. 夹道 def: position={surrounding({route|道路})}

external e.g. 外套 def: {clothing|衣物}:position={external(x: clothing|衣物)}

internal e.g. 內患 def: {mishap|劫難}:source={internal({country|國家})}

TemporalValue

TimeBefore—an event that happens after another event, e.g. 未雨綢繆 def: {prepare|準備}:TimePoint={TimeBefore({mishap|劫難})}

TimeAfter—an event that happens before another event, e.g. 始亂終棄 def: {abandon|放棄}:manner={wicked|歹}:TimePoint={TimeAfter({love|愛戀})}

Quantitative

over e.g. 出頭 def: {over()}

exact e.g. 整 def: {exact()}

approximate e.g. 左右 def: {approximate()}

PartOf—part of an object

TopPart—the top part of an object, e.g. 上半身 def: {TopPart({body|身})}

CentrePart—the center part of an object, e.g. 地心 def: {CentrePart({earth|大地})}
The E-HowNet ontology is a reconstruction of the HowNet ontology. As mentioned, it adopts the set of primitives from HowNet and follows the major type hierarchy of HowNet. The major revision was to include
the hierarchy for relations to enable semantic composition and decomposition (Chen et al., 2004). In the following, we describe the differences between E-HowNet ontology and HowNet ontology in detail.

a. Reconstruct the conceptual taxonomy of HowNet to form a single uniform taxonomy for E-HowNet:

The root of E-HowNet’s taxonomy is {all|全}. There are two subtrees, {entity|事物} and {relation|關係}, under the root. The original HowNet subtrees of {event|事件}, {entity|實體}, and {Attribute Value|屬性值} were substituted by the {event|事件}, {object|物體}, and {Attribute Value|屬性值} respectively to be subtrees of {entity|事物} of E-HowNet. The nodes of {Secondary Feature|次要特徵} and {Proper Noun|專有名詞} of HowNet no longer exist. Their subnodes are redistributed to proper position under the subtree of {object|物體}. For instance, the nodes of country names are moved to the subnodes of {country|國家}. There is no subtree of relation types in HowNet. To establish a taxonomy for relations we constructed a relational hierarchy which includes semantic roles-for-objects and roles-for-events as well as {function|函數}. The subnodes of HowNet {Attribute|屬性} were redistributed to their appropriated places under the E-HowNet subtree for {relation|關係}.

b. The major types of primitives of E-HowNet are events, objects, attribute-values, roles-for-objects, roles-for-events, and functions:

We adopt the taxonomic structures of events, objects, and attribute-values of HowNet and made some minor adjustments. The event and object subtrees are constructed by is-a (i.e. hypernym-hyponym) relations to form an inherent system. Semantic features are attached to events and objects. For instance, the argument structures are attached to different types of events. The subtree of attribute-values defines the range values of each respective attribute. The range values of an attribute are restricted to the provided attribute-values only.

c. Uniform sense representation for both content words and function words:

E-HowNet is an entity-relation model. All senses of content words, function words, and phrases are expressed by entity-relations. The semantic composition process is achieved by establishing relation between two dependent entities. Therefore E-HowNet extended the HowNet sense representation to express the relation between dependent concepts explicitly and created taxonomy for relations. For instance, in E-HowNet, the senses of function words are represented by semantic relations (Chen Y.J., Huang, Shih and Chen K.J., 2005). Both entity hierarchy and relation hierarchy are crucial for the success of complex sense representation and semantic composition process. The semantic composition and decomposition processes will be described in Section 4.

d. Revise the set of primitives. The new primitives used in E-HowNet are:
**Functions** - the kinship functions, the direction or position functions and their values, the temporal functions and values, the quantitative functions, the part-of function and their values, and the functions for expressing scopes.

**Semantic primitives for function words** - Since E-HowNet deals with senses of both content words and function words while HowNet deals with content words only, many semantic features of function words were not HowNet sememes. Therefore, many new primitives, including new features and relations were supplemented for the completion of semantic representation in E-HowNet. For instances, quantifiers of {nonreferential|無指}, {referential|有指}, {generic|通指}, {individual|專指}, {definite|定指} and {indefinite|不定指} are included. The temporal features of {SpeakingTime|說話時間}, {ReferenceTime|相對時間}, {TimeNear|時間近} and {TimeFar|時間遠} were supplemented. The referencing features of {speaker|說話者}, {listener|聽者} and {3stPerson|他人} were revised.

The senses of function words are all defined as {FuncWord|功能詞} in HowNet. In fact, function words usually mark semantic role of constituents. For instance, the preposition ‘被’ marks a agent role and its sense is defined as ‘agent={}’ in E-HowNet. Therefore, in general, the senses of function words were expressed by semantic relations. E-HowNet includes many new semantic relations not in HowNet to make the system more complete. For instance, the semantic roles of epistemic, deontic, ability, volition, and expectation are new relations for expressing senses of modal verbs and modal adverbs (Chung, Huang & Chen, 2007).

### 2.2 Principles for sense definition

#### 2.2.1 A concept is defined by its hypernym and prominent properties

Meaning of a concept is supported by its associated concepts including its formal properties, constituents, purposes, agentives, and relations to other concepts etc. To define a concept, it is not possible to encode all its associated relations. The principle for defining a concept is to first identify its immediate hypernym and then encode its most important features which suffice to differentiate it from other concepts. In principle, the qualia structure is the major representation for nominal-type (object-type) concepts (Pustejovsky, 1995), whereas event frames are for event-type concepts (Fillmore, 1998). The qualia of an object are agentive, telic, constitutive and formal. Agentive expresses the factors involved in the origin or “bringing about” of the object. Telic expresses the purpose and function of the object. Constitutive denotes the relations between the object and its constituents, such as its materials, parts, and components. Formal expresses the properties to distinguish the object within a larger domain, such as its shape, magnitude, and color. Example (11) to (14) respectively shows the usage of agentive, purpose, constitutive and formal:

(11) 早產兒 premature baby  
\[\text{def: \{嬰兒:agentive=\{labour|臨產:TimePoint=\{early|早\}:PatientProduct=\{-\}\}}\]

(12) 狗食 dog food
There are two different types of attribute features. One is simplex attribute type and another is complex relative clause type. A simplex attribute is a feature-value type and the value is expressed by some discrete elements. For instance, constitutive and formal properties can be represented by simple attribute-value pairs, i.e. Relation={Concept} pair as in the examples (11-15). A complex attribute is an eventive feature. The purpose and agenteive properties are usually represented by eventive features, which are event frames. For instances, the concepts of <老師 teacher> and <學生 student> may be defined and differentiated as <老師 teacher> def={human|人 :telic={teach|教 :agent={~}}} and <學生 student> def= {human|人 : telic={teach|教 : patient={~}}}. Event-type concepts are also defined by their hypernymous event-type, and brotherhood event-type concepts are differentiated by their frame-elements which include participant roles and adjuncts as well as their semantic restrictions. For instance, according to FrameNet II, both <request-appeal 請求> and <request-ask 要求> have the sense of <communication-request 求取>. They are differentiated by their manners:

<request-appeal 請求> Def:

{commu-reques|求取:

manner= {formal|正式}}

<request-ask 要求> Def:

{ commu-reques|求取:

manner= {informal|非正式}}

Note that the event frame and other features of <request-appeal 請求> and <request-ask 要求> are inherited from the event frame of <commu-request 求取> which has the participant roles of Speaker, Addressee, Message, and Topic according to FrameNet II.

2.2.2 Use well-defined/primitive concepts and relations to define new concepts

HowNet uses a set of primitive semantic units, called sememes, to define concepts. For example, ‘狗 dog’ is
defined as def: \{livestock|牲畜\}. Using primitives to define concepts not only causes information degrading but also fails to establish some important ontological relations between concepts. For example, HowNet defines ‘獅子狗 Beijing dog’ as def: \{livestock|牲畜\} as well, in which the hyponymy relation to ‘dog’ is missing. Thus, following HowNet, we adopt entity-relational model to define word sense. However, a concept is defined by simpler or synonymous concepts instead of semantic primitives only and all attribute relations are explicitly expressed. In E-HowNet ‘獅子狗 Beijing dog’ is defined as def: {dog|狗 source={Beijing|北京}}. With the concept ‘dog’ as the head sense, it denotes the hypernym-hyponym relation between ‘dog’ and ‘Beijing dog’, so that the definition is an ontological network in itself.

In order to achieve unambiguous and language-independent definitions, E-HowNet adopts WordNet synsets as an alternative vocabulary for conceptual indexing and representation. Take (16) for example:

(16) exhibit as evidence 證物

a. Original E-HowNet definition
   def: {physical|物質}:
   domain={police|警},
   telic= {prove|證明}:
   instrument={~}).

b. Definition is in terms of WordNet Synset id-numbers
   def: {{00010572N}:
   domain={[06093563N]},
   telic= {{00686544V+01816870V}:
   instrument= {~}}}).

c. Definition is in terms of WordNet Synset concepts
   def: {<substance>}:
   domain={<police>},
   telic= {<testify+corrobrate>:}
   instrument={~}}}).

In E-HowNet, we redefine each complex concept with its immediate hypernymous concept and major differentiation descriptions, instead of the conventional HowNet definition that uses sememes only.

e.g. site of a factory 廠址
   def: {location(工廠})
   def: {location(IndustrialPlace|場所:domain={industrial|工},telic={produce|製造}location={~})})
   def2={location([06371658N]: domain={[02579003A]},telic={[01114991V]:location={~}}})
2.2.3 Multi-level representations: High-level representations can be decomposed into primitive representations

The set of HowNet sememes (semantic primitives) are adopted by E-HowNet for ground-level definitions. In E-HowNet, new concepts are defined by any well-defined concepts and a definition can be dynamically decomposed into lower-level representations until ground-level definition is reached, in which all features in the definitions are sememes. For instance, the top level definition of ‘文學系: department of literature’ is like (17):

(17) def: {school department|學系: predication={teach|教: location={~}, content={literature|文}}}.

Since the concept ‘學系 school department’ is not a primitive concept, the above definition can be further extended into the primitive level definition (17'). The notation of ‘~’, as in HowNet, refers to the head concept of the definition which is ‘school department|學系’ in (17). Note that the feature of “predicat ion= {teach|教: location = {~}}” in (17) is redundant and will be eliminated after feature unification process (cf. section 4).

(17') def:{InstitutePlace|場所:
    domain={education|教育},
    predication={study|學習: location= {}},
    predication={teach|教: location = {}},
    predication= {teach|教: location = {~}, content={literature|文}}}.

Such a multi-level representational framework makes sense definitions more precise and easy to understand while retaining the advantage of using semantic primitives to achieve canonical sense representation.

The multilevel representation approach makes meaning representations not only more readable but also more manageable. Many basic concepts other than sememes can be used in defining new senses. For instance, <dog dog> is not a sememe, but it can be used to describe all sorts of different dogs, such as:

<great Dane 大丹狗>

def1: {狗: telic={狩獵:instrument={~},size={big|大},HumanPropensity ={gentle|柔},color={ or({black|黑},{white|白})}}}
def2: {livestock|牲畜: telic={engage|從事:content={catch|捉住:patient={animal|獸}},domain={agricultural|農},instrument={~},size={big|大},HumanPropensity={gentle|柔},color={or({black|黑},{white|白})}}}

<mast 主桅桿>

def1: {桅:telic={hang|懸掛:theme={帆},location={~}},qualification={important|重要}}
def2: {part of({ship|船}):predication={hang|懸掛:theme={PartOf({ship|船}):predication={drive|駕駛:instrument={~}},location={~}},qualification={important|重要}}

In the above two examples, the def1 uses basic concepts instead of primitives to define complex concepts. Both def1s can be discomposed into expressions in sememes as shown in def2.
Therefore, multilevel representations have the following advantages:

a. All concepts are expressed by a limited number of simple concepts.

b. More precise definitions can be achieved by using high-level concepts to define complex concepts.

c. Basic concepts are more concise for the human cognitive process.

d. Higher-level representations can be dynamically decomposed into primitive representations.

e. Higher-level representations are more readable as more information can be inherited from higher level concepts than from lower level concepts.

f. Better and easier knowledge management.

2.3 Representations for different types of senses

The sense of a natural-language sentence is the result of the composition of the senses of constituents and their relations. Lexical senses are processing units for sense composition. Conventional linguistic theories classify words into content words and function words. Content words denote entities and function words mainly mark grammatical functions. Actually, there is no clear-cut distinction between the two classes, especially for the Chinese language. In Chinese, to identify a word as a function word means it denotes more relational sense than content sense. For example, ‘被 by’ is a preposition that introduces an agent role/relation without additional content sense. On the other hand, the adverb ‘gently’ establishes a ‘manner’ relation between its content sense ‘gentle’ and the action indicated by the sentential head. By contrast, content words, such as verbs and nouns, have more content senses and less (or underspecified) relational senses. A verb denotes an event as well as the senses of its event roles. A noun refers to objects while playing the roles of verb arguments or modifiers of nouns. Therefore, it is clear that all words contain two types of senses, relation sense and content sense. The sense spectrum for syntactic categories is as shown in Table 2. For a lexical knowledge representation system, it is necessary to encode both relational senses and content senses in a uniform framework. E-HowNet is an entity-relation model to achieve representations of content/function word senses and sentence/phrasal senses. Some E-HowNet representations of word senses are shown in Table 3.

<table>
<thead>
<tr>
<th>Function words</th>
<th>Content words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relational senses</td>
<td>Content senses</td>
</tr>
<tr>
<td>de, prepositions, conjunctions, adverbs, ……………………, adjectives, verbs, nouns</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. The sense spectrum for syntactic categories

<table>
<thead>
<tr>
<th>Word</th>
<th>POS</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>因為</td>
<td>Cb</td>
<td>cause ={}</td>
</tr>
</tbody>
</table>
Table 3. Examples of E-HowNet lexical sense representations

| 下雨 | Va  | {rain|下雨} |
|------|-----|------------|
| 衣服 | Na  | {clothing|衣物} |
| 都   | Da  | manner={complete|整} |
| 潮   | Vh  | {wet|濕} |
| 了   | Ta  | aspect={Vachieve|達成} |

2.3.1 Content senses

Generally, a content word is defined by its hypernymous concept and characterization features. However some concepts with content sense do not have natural hypernymous concept. For example, the concept ‘foot’ does not have a hypernym but is defined by the concept ‘animal’ as the two concepts form a part-whole relationship. Some relations, such as kinship relations (e.g. grandfather) and directions (e.g. east) are not suitable to be defined by their hypernyms. In the following, we will illustrate the definitions of different types of content words.

Words associated by part-whole relations

To define a part, we use the ‘part of’ function and the ‘telic’ object role of the part and/or the ‘position’ function (place of the part).

腳

def: {PartOf({Animal|動物}):telic={walk|走}:instrument={~}}

Generic concepts vs. instances:

Generally speaking, the representational distinction between a generic concept and an instance is by certain features or values. For instances, {generic|通指}, which is a value of the feature ‘quantifier,’ indicates instances of generic objects, whereas other values of the feature indicates instances. The features that are hyponyms of ‘TimePoint’ or ‘location’ indicate instances of events.

凡事

def: {fact|事情:quantifier={generic|通指}}

台北區

def: {place|地方:quantifier={definite|定指},name={“台北”},position={north({Taiwan|台灣})}}

中立區

def: {place|地方:predication={not({engage|從事}):content={fight|爭鬥},location={~}}}

Words expressed by kinship relations

姑媽

def: {姑姑:qualification={已婚}}
Extension:  
\[
\text{def: } \{ \text{sister}( \{ \text{father}( \{ x: \text{human|人} \} ) \}) : \text{qualification} = \{ \text{married|已婚} \} \}
\]

表兄弟  
\[
\text{def: } \{ \text{or}( \{ \text{son}( \{ \text{sibling}( \{ \text{mother}( \{ x: \text{human|人} \} ) \} ) \} ) \}, \{ \text{son}( \{ \text{sister}( \{ \text{father}( \{ x: \text{human|人} \} ) \} ) \} ) \} ) \}
\]

Temporal related words

初戀  
\[
\text{def: } \{ \text{fact|事情}: \text{CoEvent} = \{ \text{love|愛戀} \}, \text{ordinal} = \{ \text{first|首次} \} \}
\]

決勝  
\[
\text{def: } \{ \text{HaveContest|較量}: \text{ordinal} = \{ \text{last|最後} \} \}
\]

夜空  
\[
\text{def: } \{ \text{sky|空域}: \text{duration} = \{ \text{night|夜} \} \}
\]

夜行  
\[
\text{def: } \{ \text{AlterLocation|變空間位置}: \text{duration} = \{ \text{night|夜} \} \}
\]

古籍  
\[
\text{def: } \{ \text{publications|書刊}: \text{TimePoint} = \{ \text{TimeBefore}( \{ \text{SpeakingTime|說話時間} \} ) \}, \text{TimeFeature} = \{ \text{TimeFar|時間遠} \} \}
\]

命在旦夕  
\[
\text{def: } \{ \text{die|死}: \text{TimePoint} = \{ \text{TimeAfter}( \{ \text{SpeakingTime|說話時間} \} ) \}, \text{TimeFeature} = \{ \text{TimeNear|時間近} \} \}
\]

夜貓  
\[
\text{def: } \{ \text{human|人}: \text{predication} = \{ \text{sleep|睡}: \text{agent} = \{ - \}, \text{TimePoint} = \{ \text{late|遲} \} \} \}
\]

驟至  
\[
\text{def: } \{ \text{arrive|到達}: \text{manner} = \{ \text{sudden|驟然} \} \}
\]

日久生情  
\[
\text{def: } \{ \text{love|愛戀}: \text{cause} = \{ \text{associate|交往}: \text{duration} = \{ \text{TimeLong|長時間} \} \} \}
\]

千載難逢  
\[
\text{def: } \{ \text{happen|發生}: \text{frequency} = \{ \text{rarely|偶爾} \} \}
\]

月底  
\[
\text{def: } \{ \text{month|月}: \text{TimeFeature} = \{ \text{ending|末} \} \}
\]

更年期  
\[
\text{def: } \{ \text{time|時間}: \text{duration} = \{ \text{TimeAfter}( \{ \text{suitable|適宜}: \text{theme} = \{ \text{reproduce|生} \} \) \} \}
\]

始亂終棄  
\[
\text{def: } \{ \text{abandon|放棄}: \text{manner} = \{ \text{wicked|歹} \}, \text{TimePoint} = \{ \text{TimeAfter}( \{ \text{love|愛戀} \} ) \} \}
\]

Spatial concepts (Shih et al., 2005)

Place nouns

specific place

台北  
\[
\text{def: } \{ \text{city|城市}: \text{quantifier} = \{ \text{definite|定指} \}, \text{name} = \{ \text{“台北”} \}, \text{position} = \{ \text{north}( \{ \text{Taiwan|台灣} \} ) \} \}
\]

general place

書房  
\[
\text{def: } \{ \text{room|房間}: \text{telic} = \{ \text{study|學習}: \text{location} = \{ - \} \} \}
\]
郵局  def: {InstitutePlace|場所:telic={post|郵寄:location={~}}}
橢圓  def:shape={round|方:shape={flat|扁}}
長方  def:shape={square|長：length={LengthLong|長}}
三角  def:shape={angular|角:quantity={3}}

Negation expression

Negative Polarity → not

跨黨  def: {not({distinguish|分辨}:content={黨派}}

Negative Similarity → NotSo

不遠  def: {NotSo({far|遠})}
人不可貌相  def: {NotSo({important|重要}:theme={appearance({human|人})})}
沉不住氣  def: {NotSo({calm|鎮靜})}

Verb-result compounds

We use co-index variable to indicate the difference of subject control or object control of the results.

Object-control verb-result compounds

弄壞  def: {弄:content={x},result={OutOfOrder|壞掉:theme={x}}}  
讀錯  def: {read|讀:content={x},result={misunderstand|誤信:content={x}}}

Subject-control verb-result compounds

走累  def: {tired|疲乏:cause={walk|走}}
流過  def: {GoThrough|經過:means={flow|流}}
癱坐  def: {sit|坐蹲:cause={paralyse|癱麻}}
帰納到  def: {classify|分類:content={x},cause={classify|分類:content={information|訊息}}}
併為  def: {become|成爲:theme={x},cause={merge|合併:theme={x}}}
爬起來  def: {craw}{爬:theme={x},result={arise|起身:theme={x}}}
轉過去  def: {turn|扭轉:theme={x},result={leave|離開:theme={x},distance={far|遠}}}
滾出去  def: {roll|滾:theme={x},result={leave|離開:theme={x}}}

Causative expression

病原  def: {cause|原因:predication={ill|病態:cause={~}}}
惹禍上身  def: {damage|損害:means={incur|招惹:agent={self()}}}
掃把星  def: {human|人:predication={unfortunate|不幸:cause={~}}}
Aspectual expression

Perfective

吃完  def: {eat|吃:aspect={Vachieve|達成}}

抬起來  def: {lift|提升:aspect={Vachieve|達成}}

制定出來  def: {forming|形成:aspect={Vachieve|達成}}

扣好  def: {fasten|拴連:aspect={Vachieve|達成}}

追趕上  def: {chase|追趕:aspect={Vachieve|達成}}

捉到  def: {catch|捉住:aspect={Vachieve|達成}}

Durative

瞞著  def: {HideTruth|瞞:aspect={Vgoingon|進展}}

Experiential

讀過  def: {read|讀:aspect={Vachieve|達成}}

Delimitative

楞了楞  def: {stupefied|木然;duration={TimeShort|短時間}}

數數看  def: {try|嘗試;content={count|計數}}

Pragmatic expression

尊夫人  def: {wife|{listener|聽者};attitude({speaker|說話者})={respect|敬佩}}

賤內  def: {wife|{speaker|說話者};attitude({speaker|說話者})={modest|謙}}

放屁  def: {TalkNonsense|瞎說;attitude({speaker|說話者})={ExpressAgainst|譴責}}

好漢不吃眼前虧  def: {surrender|屈服;attitude({speaker|說話者})={persuade|勸說}}

Proper noun

We use quantifier={definite|定指} and/or name="name string" to indicate a proper noun.

七星山  def: {山;quantifier={definite|定指};name="七星山";location={Taiwan|台灣}}
2.3.2 Relational senses

Function words, such as adverbs, prepositions, conjunctions, contain less content senses, but have rich relational senses. In representing the meaning of these words, we need information other than part-of-speeches because part-of-speeches do not provide the semantic information required for the unification processes for semantic composition. To make the process possible, we define function words by their relational senses and content senses (Chen, Y.J., Huang, Shih & Chen K.J., 2005). For instance, the adverb <in public|當眾> is defined as manner={overt|公開} and the preposition <by|被> is defined as agent={} with empty content. In the following, we illustrate how different types of function words are defined.

Modal words

There are five different types of modalities, i.e. epistemic, demotic, ability, volition, and expectation, in our system (Chung, Huang & Chen, 2007). Their values and relations are shown in Table 4:
<table>
<thead>
<tr>
<th>strength</th>
<th>negative/positive</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>epistemic</td>
<td>possibility</td>
<td>也許 ‘maybe’</td>
</tr>
<tr>
<td></td>
<td>impossibility</td>
<td>未必 ‘maybe not’</td>
</tr>
<tr>
<td></td>
<td>possibility</td>
<td>絕對 ‘absolutely’</td>
</tr>
<tr>
<td></td>
<td>impossibility</td>
<td>不可能 ‘impossible’</td>
</tr>
<tr>
<td>deontic</td>
<td>allowance</td>
<td>可以 ‘may’</td>
</tr>
<tr>
<td></td>
<td>disallowance</td>
<td>不用 ‘do not have to’</td>
</tr>
<tr>
<td></td>
<td>allowance</td>
<td>必須 ‘must’</td>
</tr>
<tr>
<td></td>
<td>disallowance</td>
<td>不應該 ‘be not supposed to’</td>
</tr>
<tr>
<td>ability</td>
<td>capacity</td>
<td>會 ‘can’</td>
</tr>
<tr>
<td></td>
<td>incapacity</td>
<td>不克 ‘not really can’</td>
</tr>
<tr>
<td></td>
<td>capacity</td>
<td>能 ‘be able to’</td>
</tr>
<tr>
<td></td>
<td>incapacity</td>
<td>不會 ‘cannot’</td>
</tr>
<tr>
<td>volition</td>
<td>willingness</td>
<td>想 ‘hope to’</td>
</tr>
<tr>
<td></td>
<td>unwillingness</td>
<td>不想 ‘do not hope to’</td>
</tr>
<tr>
<td></td>
<td>willingness</td>
<td>要 ‘want to’</td>
</tr>
<tr>
<td></td>
<td>unwillingness</td>
<td>不要 ‘do not want to’</td>
</tr>
<tr>
<td>expectation</td>
<td>unexpectedness</td>
<td>果真 ‘really’</td>
</tr>
<tr>
<td></td>
<td>unexpectedness</td>
<td>不料 ‘unexpectedly’</td>
</tr>
<tr>
<td></td>
<td>unexpectedness</td>
<td>果然 ‘as expected’</td>
</tr>
<tr>
<td></td>
<td>unexpectedness</td>
<td>竟 ‘very unexpectedly’</td>
</tr>
</tbody>
</table>

Table 4. E-HowNet semantic relations and values for modal senses

也許 def: possibility={ish|稍}
信得過 def: {believe|相信:capacity={ish|稍}}
毋遺後患 def: {PassOn|留給:possession={mishap|劫難},disallowance={extreme|極}}
他不可能來 def: {come|來:theme={3rdPerson|他人:gender={male|男}}, impossibility={extreme|極}}
你不必來 def: {come|來:theme={listener|聽者},disallowance={ish|稍}}

Conjunctions

Conjunctions are function words marking semantic relations between two constituents. The conjunctive relations and respective conjunctive words are shown in the following hierarchy (Figure 2):

e.g. 因為 def: cause={}
e.g. 所以 def: result={}
Figure 2. Taxonomy of logical relations including conjunctions

**Adverb**

Adverbs have partial relational sense and partial content sense. For example:

- e.g. 透頂   def: degree={very|很}

  西瓜冰涼透頂   def: {cool|凉:degree={very|很},theme={watermelon|西瓜}}

**Prepositions**

Usually a preposition marks different semantic roles. Hence it is ambiguous and has multiple definitions.

  向   def: source={} 台北向印尼購買天然氣
def: target={} 本黨鄭重向全台灣人民宣布
def: direction={} 而車子還是不知道開向何方

藉由 def: instrument={object|物體} 藉由此蒸氣孔釋出多餘的能量
def: means={event|事件} 希望藉由外派，讓公司幫他辦移民

隔 def: TimePoint={TimeAfter()} 五零年代的《生死戀》，是華人女性隔了二十年再次擔任主角
def: {from|相距:location={}} 秦軍和晉軍隔着肥水遙遙相對

Question words

為何 def: reason={Ques()}

誰 def: participant={human|人: formal={Ques()}}

怎麼 def: means={Ques()}

Non-predicative Adjective

瓶裝 def: predication={ wrap|包紮:instrument={瓶},patient={~}}

萬用 def: qualification={ role({thing|萬物}):kind={various|多種}}

商用 def: predication={use|利用:domain={economy|經濟},instrument={~}}

藥用 def: predication={use|利用:telic={doctor|醫治:instrument={~}}}

歐式 def: source={Europe|歐洲}

祖傳 def: source={forefathers|祖先}

棉製 def: material={棉}

無菌 def: predication={not({own|有}):agent={~},possession={bacteria|微生物}}

植物性 def: source={plant|植物}

備用 def: predication={SetAside|留存:telic={replace|代替},theme={~}}

宮殿式 def: predication={alike|似:contrast={object|物體:source={宮殿}},theme={~}}

小寫 def: qualification={pattern({character|文字}):qualification={easy|易}}

Determinatives

3 Although non-predicative adjectives are content words, they are discussed along with functional categories as they are represented by relational sememes.
Demonstrative determinatives

這 def: quantifier={definite|指}

第 def: ordinal={}

Specifying determinatives

每 def: range={all|全}

其他 def: qualification={other|另}

Numeral determinatives

一 def: quantity={1}

Quantitative determinatives

許多 def: quantity={many|多}

有的 def: quantity={some|些}

之多 def: approximate()

Interrogative determinatives (Huang & Chen, 2008)

啥 def: formal={Ques()}

Measure words (Tai et al., 2009)

Measure words with content sense:

碗 def: container={bowl|碗}

米 def: length={meter|公尺}

月 def: duration={month|月}

樣 def: {kind({object|物體})}

Measure words without content sense:

本 def: null

間 def: null

宗 def: null

2.4 Syntax of E-HowNet expressions

The syntax of E-HowNet expressions (sense representations) follows a set of formal syntax rules (see
appendix A). The basic tokens of E-HowNet expressions are: concepts, relations, functions, variables, constants, and symbols. Concepts are sememes and complex concepts expressed by E-HowNet expressions. Relations are semantic roles. Functions are members of \{\text{not}(), \text{and}(), \text{or}(), \text{not-so}(), \text{part-of}(), \text{Ques}(), \text{father}(), \text{east}(), \ldots\}. Variables are \{\sim, X, Y, Z, \ldots\} where \“\sim\” denotes the head concept of the expression and variable X, Y, Z are for co-indexing. The constants refer to some particular instances, such as name strings (e.g. ‘台北’) and individual constants, such as \{\text{Speaker, Listener, SpeakingTime} \ldots\}. The following symbols \{‘:’, ‘=’, ‘(’, ‘)’, ‘{’, ‘}’, ‘,’\} are delimiters for E-HowNet expressions. Table 5 shows the basic expressions of E-HowNet.

<table>
<thead>
<tr>
<th>Concept def:=</th>
</tr>
</thead>
<tbody>
<tr>
<td>{Hypernym : Feature,\ldots, Feature}, or {Concept} or {Sememe};</td>
</tr>
<tr>
<td>The expression means that a concept may be defined by 1) its hypernymous concept and semantic features, or 2) a synonymous concept, or 3) a primitive concept.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Features def:=</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relation(x)={Concept};</td>
</tr>
<tr>
<td>The expression says that a semantic feature is expressed by a (Relation, Concept) pair, which denotes the semantic relation (Relation) between semantic feature (Concept) and the argument x. Arguments are in the range of {\sim, Speaker, Listener,\ldots}. Relation(\sim)={Concept} will be abbreviated as Relation= {Concept}.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relation def:=</th>
</tr>
</thead>
<tbody>
<tr>
<td>property, content, host, location, agent, patient,\ldots;</td>
</tr>
<tr>
<td>a set of semantic relations.</td>
</tr>
</tbody>
</table>

**Table 5. Syntax of basic E-HowNet expressions**

The detailed syntax rules are shown in Appendix A.
3. Advantages of E-HowNet

The E-HowNet intends to bridge the gaps between string processing and conceptual processing. It has the following advantages in semantic processing.

a) Sense representations are precise and incremental:
   e.g.  
   
   <great dane|大丹狗> Def:=  
   {dog|狗:  
   location={German|德國},  
   telic={hunt|狩獵:  
   instrument={~},  
   size={big|大型},  
   HumanPropensity={gentle|溫和},  
   color={black white|黑白}}

   A pure taxonomy approach, such as WordNet, does not provide detailed description of a concept.

b) Conceptual classes are characterized by features.

For example, <great dane> is also classified as <hunting instruments> and <animal with black and white colors> according to its telicity feature and color feature respectively. Other examples of elements of the class of hunting instruments are:

<hunting gun|獵槍> def:  
{gun|槍:  
   telic={hunt|狩獵:  
   instrument={~}}}

<trap|陷阱> def:  
{facility|設施:  
   telic={hunt|狩獵:  
   instrument={~}}}

Although there is no natural class called <Animals with black and white colors>, such a class can be described by the feature set of {beast|走獸: color={black white|黑白}} which happens to be the features shared by the following examples:

<panda|貓熊> def:  
{beast|走獸:  
   location={China|中國},  
   predication={eat|吃:  
   patient={bamboo|竹子},  
   agent={~},  
   color={or({black|黑},{white|白})}}

<zebra|斑馬> def:  
{horse|馬:
c) Achieves near canonical semantic representation:

If two sentences have the same meaning but different surface forms or in different languages, may have similar E-HowNet representations. For example:

(1) 我 買了 一本 科幻小說。
(2) I bought a science fiction.

Both sentences have the same representation of {buy: agent={speaker|說話者}, possession={小說: qualification={or({scientific|科學},{fake|僞})}, quantity={1}}, TimePoint={TimeBefore({Speaking time|說話時間})}}.

Note that the above high-level representation can be extended to ground level and/or WordNet synset representations.

d) Multi-level meaning representations through semantic decomposition:

A semantic expression can be defined by any well-defined concepts in E-HowNet which can be further decomposed into representations of primitive concepts.

e.g.  <tailor store|裁縫店>  def:= {store|店: telic={sew|裁縫: location={~}} } can be extended to:

{InstitutePlace|場所: telic={produce|製造: PatientProduct=

{clothing|衣物},
location={~}}}

By contrast, in HowNet concepts are defined by primitive concept sememes only. In the above example, the basic concept <InstitutePlace|場所> does not have the information of “commerce” inherited from <store|店>.

e) As a conceptual representation that may use WordNet synsets as its description language, E-HowNet is universal and language-independent.

E-HowNet expressions can be converted into expression of WordNet synsets like def2 below:

e.g. bulletin board|公佈欄

def1:{facilities|設施:telic={put|放置:location={~},theme={text|語文:predication={announce|發表:content={~}}}}}

def2:{{(establishment):telic={[(put,set,place)]:location={~},theme={[(text,textual_matter)+(command_language,query_language,search_language)]:predication={[(announce,denote):content={~} ]}}}}}

E-HowNet links different ontologies. For instance, we established the links between HowNet sememes and WordNet synsets. Thus WordNet synsets are used as an alternative intermediate representational language. In the future, we will link events of E-HowNet to the event frames of FrameNet.
4. Semantic Composition and Decomposition

Semantic composition and decomposition are achieved by feature unification. During the unification process, feature values of the same relation type are unified. For instance, in example (17), the hypernymous class <school department|學系> of <department of literature|文學系> is not a primitive concept and is decomposed into the definition of {InstitutePlace|場所: domain= {education|教育}, predication= {<study|學習>: location={~}}, predication= {teach|教: location={~}}}. Then, the reduplicated features of predication={teach|教:location={~}} appear after the decomposition process, as in (17'). Finally, the reduplicated features will be unified into a single feature of “predication={teach|教:location={~}, theme= {literature|文}}” and the ground level representation of <department of literature|文學系> becomes:

def:{InstitutePlace|場所:
    domain= {education|教育},
    predication= {study|學習:location= {~}},
    predication= {teach|教:location={~},content={literature|文}}}

In the semantic composition process, if two constituents are syntactically dependent, their E-HowNet representations will be unified by following the basic composition process below.

**Basic semantic composition process:**

If a constituent $B$ is a dependency-daughter of the constituent $A$, i.e. $B$ is a modifier or an argument of $A$, then unify the semantic representation of $A$ and $B$ by the following steps:

**Step 1:** Disambiguate the senses of $A$ and $B$.

**Step 2:** Identify the semantic relation between $A$ and $B$ to derive relation($A$)={$B$}.

**Step 3:** Unify the semantic representation of $A$ by inserting relation($A$)={$B$} as a sub-feature of $A$.

The methods for word sense disambiguation and relation identification are out of the scope of this manual. We will not discuss those issues here. In the following sentence, we will show how step 1 and 2 are done:

(18) *Because it was raining, the clothes are all wet.* 因為下雨，衣服都濕了。

In (18), ‘濕 wet’, ‘衣服 clothes’ and ‘下雨 rain’ are content words, whereas ‘都 all’, ‘了 le’ and ‘因爲 because’ are function words. Their E-HowNet sense representations are shown in Table 3. The main difference in their representations is that the function words are represented by relations of the form rel(x)=(y), whereas the content words do not make references to the semantic roles they involve in the definition. When a content word is a dependency daughter of a head concept, the relation between the head concept and this content word needs to be established by a parsing process. Suppose that the following dependency structure and semantic relations (19) are derived by parsing sentence (18):

Then, (20) is the semantic composition which results from the unification process. The dependency daughters become feature attributes of the sentential head ‘wet|濕.’

(20) def:{wet|濕:
    theme={clothing|衣物},
    aspect={Vachieve|達成},
    manner={complete|整},
    cause={rain|下雨}}.

In (20), the function word ‘因為 because’ links the head concept ‘濕 wet’ and ‘下雨 rain’ with the ‘cause’ relation. The result of composition is expressed as cause({wet|濕})={rain|下雨}. For the sake of notational convenience, the head argument of a relation is omitted. Therefore cause({wet|濕})={rain|下雨} is expressed as cause={rain|下雨}; theme({wet|濕})={clothing|衣物} is expressed as theme={clothing|衣物} and so on.
5. Potential Applications of E-HowNet

There is still a long way to go to achieve fully automatic semantic composition and natural language understanding. Many research problems and difficulties need to be resolved, such as robust syntactic parsing, word sense disambiguation, unknown word identification, semantic role assignment, semantic composition, aspectual normalization, and canonical sense representation etc. Such technologies are indispensible tools and hot research topics for NLP (Tai et al., 2009; Shih et al., 2006; Chen & Chen 2000; Bai & Chen 1998). E-HowNet does not provide the solutions for the above problems directly but it provides a valuable resource in solving those problems. Other than semantic generalization and specialization, some specific applications of E-HowNet are exemplified below.

5.1 Identify senses of new compound words

Veale (2005) tests the ability of HowNet system in doing analogy generation and concludes that HowNet contains sufficient structure to realistically support both a taxonomic abstraction view and a structure-mapping view of analogy generation. Since E-HowNet adopts and extends the sense definition mechanism of HowNet, we can use similar strategy to discover the semantic structures of a very productive type of unknown words, for instance compound nouns.

E-HowNet uses hypernymous concepts to classify concepts and differentiates concepts of same hypernymous class by their major features (Shih et al., 2006). To discover the sense and semantic structure of a noun compound is to disambiguate the semantic ambiguity of the morphological head of a compound noun and find the proper semantic relation between constituents of the compound. For example, when we see the unknown/undefined compounds such as 牧工 “hired herdsman”, 核工 “nuclear industry”, or 唱工 “art of singing”, firstly, we have to find the appropriate meaning for each head of these unknown compound. Secondly, we have to build the correct relation between their modifiers and the heads, such as the relation between 牧 and 工, 核 and 工, etc.

Chen & Chen (2000) proposed an example-based similarity measure to disambiguate the polysemous heads. They extracted some examples with the polysemous head morpheme from corpora and dictionaries, and classified them into different groups according to their meaning. Let’s take “工” as example and add E-HowNet definitions for each class, shown as Table 6.

The meaning of 牧工 “herdsman”, 核工 “nuclear industry”, or 唱工 “art of singing” are then determined by comparing the similarity between their modifiers and the modifiers of each class of examples. That is, we compare 牧, 核 and 唱 respectively with 搬運, 女, 童, 化, 機, 刀...etc. And then find the most similar examples and choose their semantic type as the semantic type of the target words. For instance, 牧 is most similar to the modifiers in first class, thus the semantic type of 牧工 is “labor”. Similarity calculation helps to work out a preliminary definition for each unknown/undefined compound. To further define them, we need to know the relation between the modifiers and their head. Suppose that all examples in class two are shared with the same semantic feature “domain”, then we can further define 核工 by replacing the value of feature “domain” with the sense of “核” to create a new definition as (21):

(21) 核工 “nuclear industry”
In similar way, 唱工 can be defined as (22):

(22) 唱工 “art of singing”
def:{skill|技術:telic={sing|唱:method={~}}}
The sense of \(<\text{從}\>\text{from}\) can be disambiguated by the respective semantic restrictions either \(<\text{place}\>\) or \(<\text{time}\>\) of its argument.

### 5.3 Semantic role assignment

The problem of semantic role assignment is a hot research topic. In E-HowNet, ample conceptual relations are encoded in the lexical sense representation, providing a knowledge base for identifying semantic relations between two concepts (cf. section 5.1). In addition, all event frames including argument roles are provided at event hierarchy of E-HowNet.

Some semantic relations are indirect and hard to identify. For instance, the relations between \(<\text{fast}\>\) and \(<\text{food}\>\) between \(<\text{fast}\>\) and \(<\text{car}\>\) are different and cannot simply be described as property-entity relation. The semantic gaps regarding \(<\text{serving fast}\>\) and \(<\text{moving fast}\>\) respect to two compounds are not expressed explicitly. The different telic feature values for \(<\text{food}\>\) and \(<\text{car}\>\) may provide some clues to resolve the problem. We will elaborate the problem more in the next section.

### 5.4 Filling semantic gaps by automatic deduction

In real implementations of semantic composition, we have found filling semantic gaps an important task, because some semantic elements are frequently omitted from surface sentences. To that end, we have encoded event frames and construction patterns to the respective verbs and keywords in the E-HowNet system. We have not only established object-attribute relations, but also revealed the participant roles in an event. For instance, ‘color’ is a semantic role that denotes the relation between an object x and its color range y, as expressed by color(x)={y}. In the following sentences (23)-(25), we demonstrate how to restore sense omissions by object-attribute relations.

(23) I like the red something 我喜歡紅的

\[
\text{def: \{FondOf|喜歡:}
\]
\[
\begin{align*}
\text{experiencer} &= \{\text{speaker|說話者}\}, \\
\text{target} &= \{\text{object|物體:}\}, \\
\text{color} &= \{\text{red|紅}\}\}
\end{align*}
\]

Because the semantic role ‘color’ is an attribute of objects, it implies an object was missing in the sentence (23) and thus it is known that the target of ‘like’ has to be recovered from context. Similarly, Quantitative Determinative is a semantic role that establishes the relation between an object and its quantity. A representation like that in (24) thus signifies the presence of an object. For the same reason as exemplified in (23), we know the object is omitted in (25) too. The event frame of \{\text{speak|說}\} has been coded as to take \{\text{human|人}\} as an agent role. We therefore know the absent object has to be an instance of \{\text{human|人}\}.

(24) few 少數

\[
\text{def: quantity={few|少}.}
\]
(25) There are only a few who dare to speak out. 敢說話的是少數

def:{dare|敢於|content={speak|說}, experiencer={human|人}, quantity={few|少}}.

By the same token, we can figure out what are semantic relations between {fast|快} and <food 餐> and <car 車> in 快餐 and 快車. Since {fast|快} is a value of the event-attribute ‘speed,’ it has to modify events rather than objects such as 餐 or 車 and the feature ‘speed’ is most likely associated with the telic features of 餐 and 車, i.e. <serve > and <move> respectively.

The way to fill semantic gaps of constructions is by providing a mapping table to connect the grammatical functions and fine-grained semantic roles (Huang, Shih and Chen, 2008). The most typical example is the comparative construction for ‘比’ bi’. The sense of ‘bi’ comprises a complex argument structure which is shown in (26). Sentence (27) is its implementation:

(26) ‘bi 比’ def: contrast={} in the course-grained event frame of {AttributeValue: theme={}, contrast={}, quantity (or degree)={}, manner={}, location={},time={}}.

(27) I am taller than him by a head.我比他高一個頭

Surface structure: theme[NP]+contrast[PP[比]]+Head[V]+quantity

Parsing result: {tall|高:

theme={speaker|說話者},
contrast={他},
ComparativeQuantity={head|頭:quantity={1}}).

The grammatical roles and the thematic roles can be automatically extracted from a sentence (You & Chen, 2004). Then, through a mapping table that connects the grammatical functions and fine-grained semantic roles, the machine is able to identify the thematic role “theme” and “contrast” refer to the fine-grained semantic roles Profiled_Item+Profiled_Attribute and Standard_Item+Standard_Attribute and that the Profiled and Standard+Attributes need to be restored. The Head ‘高’ suggests that the attribute to be restored should be 身高. (28) shows the semantic representation with the semantic gap filled in:

<table>
<thead>
<tr>
<th>Fine-grained Semantic Roles</th>
<th>Thematic Roles</th>
<th>Grammatical Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profiled_Item+(Profiled_Attribute)</td>
<td>Theme; Experiencer</td>
<td>Subject</td>
</tr>
<tr>
<td>Standard_Item+(Standard_Attribute)</td>
<td>Contrast</td>
<td>Object[PP[bi]]</td>
</tr>
<tr>
<td>Comparison_set</td>
<td>Head</td>
<td>Verb</td>
</tr>
<tr>
<td>Attribute_Value</td>
<td>ComparativeQuantity;</td>
<td>Complement</td>
</tr>
<tr>
<td>Degree</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7. Mapping table for the fine-grained semantic roles

<table>
<thead>
<tr>
<th>Degree</th>
<th>Manner</th>
<th>Adjunct (Manner)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manner</td>
<td>Manner</td>
<td>Adjunct (Location)</td>
</tr>
<tr>
<td>Place</td>
<td>Location</td>
<td>Adjunct (Time)</td>
</tr>
<tr>
<td>Time</td>
<td>Time</td>
<td>Adjunct (Time)</td>
</tr>
</tbody>
</table>

(28) My height is one head taller than his height. 我的身高比他的身高高一個頭

\[
\text{def: \{} \text{tall|高:} \\
\text{Profiled\textunderscore Item=} \{\text{speaker|說話者}\}, \\
\text{Profiled\textunderscore Attribute=} \{\text{height|身高}\}, \\
\text{Standard\textunderscore Item=} \{\text{他}\}, \\
\text{Standard\textunderscore Attribute=} \{\text{height|身高}\}, \\
\text{Degree=} \{\text{head|頭;quantity=} \{1\}\} \}. \\
\]

5.5 Toward near-canonical meaning representation

Through semantic composition process we can derive semantic representations of phrases as well as sentences. In addition, E-HowNet sense representation is a conceptual representation which is language independent and near-canonical. For instance, two sentences of similar meaning but with different surface forms may derive similar E-HowNet representations.

e.g. 機長機敏地抓獲女搶犯 vs. 飛機駕駛員敏捷的逮捕女強盜

After syntactic parsing, the event structures of two sentences are:

\[
\text{def: \{} \text{抓獲} : \text{agent=} \{\text{機長}\}, \text{patient=} \{\text{搶犯;gender=} \{\text{女}\}\}, \text{manner=} \{\text{機敏}\} \} \\
\]

vs.

\[
\text{def: \{} \text{逮捕} : \text{agent=} \{\text{飛機正駕駛}\}, \text{patient=} \{\text{強盜;gender=} \{\text{女}\}\}, \text{manner=} \{\text{敏捷}\} \} \\
\]

The above two event structures apply decomposition process, and then derive similar results as shown below.

\[
\text{def: \{} \text{catch|捉住:} \\
\text{agent=} \{\text{human|人;predication=} \{\text{manage|管理;agent=} \{\sim\}, \text{patient=} \{\text{aircraft|飛行器}\}\}\}, \\
\text{patient=} \{\text{human|人}; \text{HumanPropensity=} \{\text{guilty|有罪}\}, \text{predication=} \{\text{rob|搶}; \text{agent=} \{\sim\}, \text{gender=} \{\text{female|女}\}\}, \text{manner=} \{\text{clever|聰}\} \} \\
\]

vs.
def: {catch|捉住}:agent={human|人}:predication={manage|管理}:agent={~,patient={aircraft|飛行器}}},
  patient={human|人}:HumanPropensity={guilty|有罪},
  predication={rob|搶}:agent={~, gender={female|女}}, manner={nimble|捷}}

Nevertheless, true canonical representation is not yet achieved. To discover different aspects of similar events needs normalization of sense representations. For instances, <buy> and <sell> are typical examples of the same event from different viewpoints. Should they normalize to the same semantic representation?
6. Conclusions and Future Research

HowNet proposed a new model to represent lexical knowledge, inspiring us to expand this framework to achieve the task of mechanical natural language understanding. E-HowNet confines each concept to a semantic type and defines the relation between these types. Hence we have a consistent approach to representing concepts so that the computer can process and relate meanings.

Semantic composition is a crucial component of language understanding. We have proposed a uniform representation system for both function words and content words to achieve semantic composition. We also suggested compositional functions to extend the expression of new concepts and make word and phrase definitions more detailed and accurate. Since sense omission increases the potential for wrong readings, we try to fill semantic gaps by automatic inference through the framework of E-HowNet.

There are still many obstacles to achieving the goal of automatically extracting knowledge from language. Apart from sense disambiguation, discord between syntactic structures and their associated semantic representations is another critical problem. We need to determine rules which map from coarse syntactic structures to fine-grained semantic relations. Gap filling processes, as discussed, need to be an integral part of the mechanism. Normalization of sense representation to achieve real canonical sense representation and fine-grained semantic representations are also indispensable. Our future research will continue to address these issues.
References


Appendix A. Formal syntax of E-HowNet Expressions:

Concept → Complex-Concept | Basic-Concept
Complex-Concept → '{' (Basic-Concept | Co-indexed-Concept) ': Feature-Values '}'
    → '{' (Basic-Concept | Co-indexed-Concept) '}'
Basic-Concept → Sememe | Mapped-Concept | Intermediate-Form | Variable | Constant
Co-indexed-Concept → Basic-Concept '=' Variable
Variable → ℝ | 'X' | 'Y' | 'Z'
Constant → CC* ;
CC→ Chinese-Character;
Mapped-Concept → Relation (' Concept ') | Function (' Mapped-Concept ');
Concepts → Concept | Concepts ', Concept | Null;
Chinese-Word → CC* ;
Feature-Values → Feature-Value | Feature-Values ', Feature-Value;
Feature-Value → Feature '=' Complex-Concept
Feature → Mapped-Concept | Relation;

Note: The set of sememes and relations can be accessed from “E-HowNet Ontology Online” at http://ckip.iis.sinica.edu.tw/taxonomy/.