詞　庫　小　組

技術報告　09-01

Technical Report no. 09-01

Lexical Semantic Representation and Semantic Composition
An Introduction to E-HowNet

Ⓒ詞庫小組　2009年9月

出版處：台北, 南港
中央研究院　資訊科學研究所
中文詞知識庫小組

Lexical Semantic Representation and Semantic Composition

An Introduction to E-HowNet

Lexical Knowledge Representation and Semantic Composition

--An introduction to E-HowNet

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 (basically basic concepts) and conceptual relations.
b. A uniform sense representation schema 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, Yu-Ming Hsieh who contributed to the development and design of E-HowNet.

E-HowNet Version 2.0

The current E-HowNet ontology shown on the web is the result of automatic constructed by a computer program according to the pre-defined hierarchical structure of primitive and basic concepts as well as E-HowNet expressions for all word entries.

The major improvements of E-HowNet version 2.0 are:

a) Reorganizing the hierarchical structure of primitive and basic concepts: We extend a large set of basic concepts which make a deeper hierarchical structure and more precise semantic branching. It also results that lexical senses expressed based on basic concepts became more precise and readable. We also adjust the ontology structure into two parts. The first part is hierarchy for entities and the second part is hierarchy for relations, i.e. semantic roles. Furthermore the Attribute types and Value types are correspondently organized.
b) Rich lexical information: In addition to sense definition, each entry of lexical sense may also include operational expressions as well as semantic functions which facilitate future semantic composition processes. Event frames, i.e. argument structures, of event type primitives are also provided.
c) Developing a new automatic ontology reconstruction system: In case of revisions of lexical sense expressions or nodes of conceptual hierarchy, the ontology reconstruction system may re-attach each lexical entry to appropriated ontological nodes and results a new ontology.
d) Improvement of sense definitions and sense definitions for basic concepts: Many word sense definitions are revised and became more precise and readable by using basic concepts in their sense expressions. More semantic links are established due to shared semantic features as well as explicit relation links, such as antonym, attribute-value, entailment etc.

Wei-Yun Ma and Keh-Jiann Chen
CKIP group http://rocling.iis.sinica.edu.tw/CKIP/
Institute of Information Science, Academia Sinica
July 2016

The 2016.7Revision

Lexical Semantic Representation and Semantic Composition

An Introduction to E-HowNet

/* Please check the format, labels, and numbers of text, examples, figures, and tables.

1. 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-based 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: (a) word similarity (劉 & 李, 2002), (b) machine translation and (c) 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)

(1) Word : sense1 : grammatical function   semantic knowledge   world knowledge
sense2: grammatical function   semantic knowledge   world knowledge
sense3: …

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:

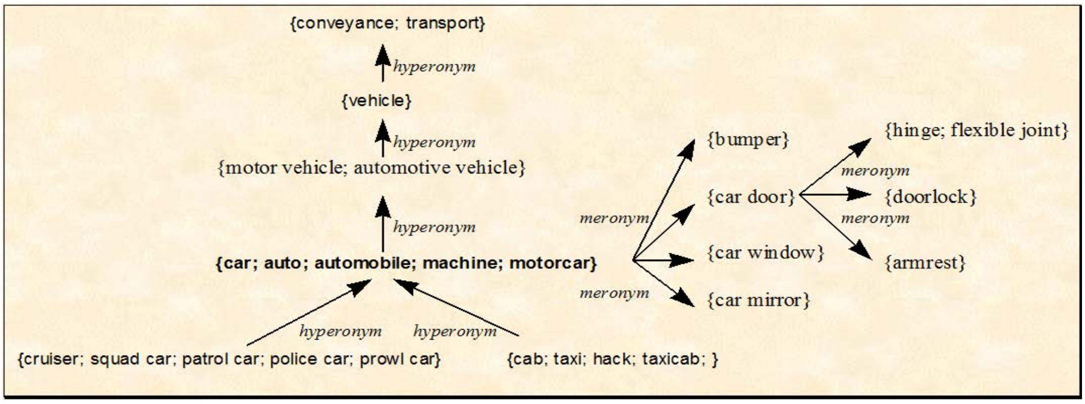

Table1. Synset :{car; auto; automobile; motorcar}

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 cause 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 (2) :

(2) {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	V2.0AlterRelation|變關係
V1.01 isa|是非關係	V2.01 AlterIsa| 變是非
…	
V1.1state|狀態	V2.1AlterState|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/).

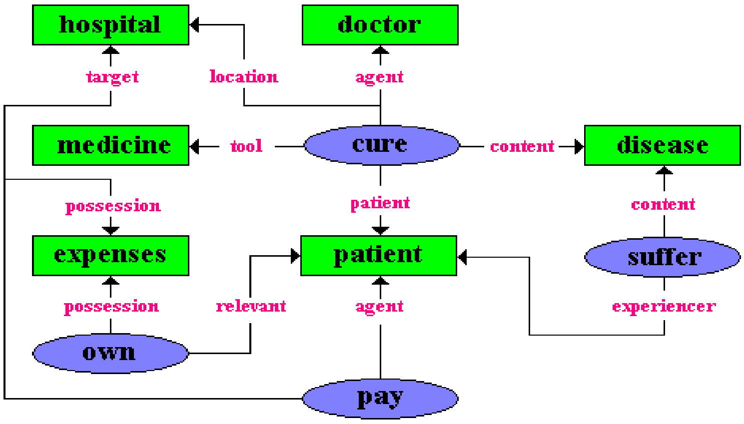

Figure 1. An example of the lexical representations of HowNet

The disadvantages of the HowNet approach are:

a. Representation by primitives degrades precision and readability.

i. *tiger* 老虎 def:{beast|走獸} and *bear* 熊 def:{beast|走獸};

ii. *forceps* 鉗子 def:{tool|用具:{hold|拿:instrument={~}}};

iii. *watchmaker's shop* 鐘錶店 def={InstitutePlace|場所:{buy|買: location={~},possession={tool|用具: {tell|告訴: content={time|時間}, instrument={~}}}}, {repair|修理: location={~}, patient={tool|用具:{tell|告訴: content={time|時間}, instrument={~}}}},{sell|賣: location={~}, possession={tool|用具: {tell|告訴: content={time|時間}, instrument={~}}}}}

b. Semantic relations are not explicitly expressed.

c. Sense of function words and relational concepts are not well established. e.g. function word *just* 僅 def:{FuncWord|功能詞:emphasis={?}}

d. 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).

a. Word senses (concepts) are defined by not only primitives but also any well-defined concepts and conceptual relations. Thus phrasal senses can be similarly expressed by semantic composition and decomposition processes. (Note: In real implementation of E-HowNet 2.0, word senses are defined by primitives and basic concepts only. Phrasal senses are intended to be derived and represented automatically by any well-defined concepts.)

b. A uniform representation model for function words and content words, as well as phrases.

c. Semantic relations are explicitly expressed for all meaning representations.

d. Semantic composition and decomposition capabilities.

e. Near canonical representations for lexical senses and phrasal senses.

The E-HowNet system comprises the following components:

a. The E-HowNet ontology at http://ckip.iis.sinica.edu.tw/taxonomy/

b. A set of primitive concepts in the form of {English|中文} (called sememes in HowNet) which include events, objects, and relations

c. A set of basic concepts in the form of {中文|English} and each basic concept is defined by sememes

d. 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 and dissimilarity between two sentences or two words. To achieve the above goals, it requires the support of ontologies. Ontology provides the following functions.

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 index 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.

In E-HowNet 2.0 all concepts are either primitive concepts or defined by simpler concepts (either primitive concepts or basic 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). A primitive concept will have an English equivalent beside it, e.g. {read|讀}, whereas a basic concept will be expressed by a Chinese word and its English translation pair which is further defined by primitive concepts, e.g. {狗|dog} defined as {livestock|牲畜:telic={TakeCare|照料:patient={family|家庭},agent={~}}}.

The concepts form a hierarchical structure by is-a (hyponymy) relation, as shown in 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. Additional semantic linking was also established by other lexical information shown in the section 4.1.

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.[1] The entity subtree is formed by event subtree and object subtree. The relations include semantic-roles and logical 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). Semantic roles also form a hierarchical structure from coarse-grained semantic roles to fine-grained semantic roles.

All semantic roles are binary relations rel(x,y), with the parameter x usually being the dependency head of a constituent and y being dependent daughter. 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 values of domain and range depend on the relation type. In HowNet the ranges of attribute types of relations are their 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.

[1] See http://ckip.iis.sinica.edu.tw/taxonomy/ for details.

Relation concepts may also play the role of subject/object and consider as a subtype of object entities, called relation-entities, and have the form of expressions like entities, such as {color|顏色}, {location|地點}, {cause|原因} etc. They form a subtree of {relation|關聯} under the node of {object|物體}. {relation|關聯} subtree and relation subtree (i.e. semantic roles) are in parallel but have different senses and are different subtrees in E-HowNet ontology.

Function is a special kind of relation, i.e. a one-to-one 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|父})} denotes 'the grandfather of x' and the direction function of north{north({east({place|地方})})} 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. Different function may have different semantic types. (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({wife|妻子})}.

(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 expresses the semantics of the target object.

In E-HowNet, we also regard union, and, or relation, question and negation relation as logical 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: cause={Ques|疑問}.

(9) *be frown on* 不悅

def: {not({joyful|喜悅})}.

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. *speed of wind* 風速

def: {speed({wind|風})}

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: { electricity| 電 : length({phenomena| 現 象 :cause={shiver| 顫動:theme={ ~}}})={10 公里}}

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

2.1.2 Relations and the Usages of E-HowNet's Semantic Roles

The relation tree is formed by subtree of semantic roles and subtree of logical functions. Semantic roles are major elements of relations. There are three sub-types of semantic roles. They are roles for object, roles for attribute and roles for event. Semantic roles are organized in a hierarchical way similar to the taxonomy for entities. A hyponym role entails its hypernym role. Their usages are exemplified in the following sub-sections and the full set of semantic roles are demonstrates in Appendix A. The logical functions, such as *and*, *or*, *union*, are special kind of relations which map arguments into a unique value.

a) Semantic roles for objects

For example, *quantifier* is a major semantic role for object. The hyponym roles of *quantifier* includes *quantity, rate, amount, container and sequence.*

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}}

rate —a specific kind of ratio, in which two measurements are related to each other, e.g. 出生率 def: rate({ComeToWorld|問世})

amount — an enumerable quantity, e.g. 三個 def: amount={3}

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

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

b) Semantic roles for attributes

Host and *value* are two major semantic roles for attributes to form the basic semantic unit of attribute(host)={value} triple.

host — host of attributes, e.g. 地心引力 def: {strength({attract|吸引}):host={earth|大地}}

value — value of attributes, e.g. 台籍 def: {nationality({human|人}):value={台灣|Taiwan}}

c) Semantic roles for events

Following shows a hierarchical structure regarding semantic roles of *actor*.

actor—the actor of an event.

agent—a conscious actor which performs an action with control (on purpose) and has a physical, visible effect on object, e.g. 工讀生def: {學生|student:predication={打工|WorkPartTime:agent={~}}}

experiencer—an animate being who perceives a stimulus or registers a particular mental or emotional process or state, e.g. 好戰份子def: {human|人:predication={FondOf|喜歡:target={fight|爭鬥},experiencer={~}}}

causer—an unconscious force which incurs an event without purpose, e.g. 病媒蚊 def: {蚊子|mosquito:telic={infect|傳染:theme={disease|疾病},causer={~}}}

d) LogicalFunction

Union—union of two elements or sets into a larger set, e.g. 父女 def: {union({father|父},{daughter|女兒})}

and—juxtapose objects or events, e.g. 跟 def: and(); 又驚又喜 def:{and({joyful|喜悅},{surprise|驚奇})}

or—the concepts which are alternatives, e.g. 藍白色 def: color={or({blue|藍},{white|白})}; 出入 def: {or({GoInto|進入},{GoOut|出去}) };

not—negates an event, e.g. 不理不睬 def: {not({ShowInterest|理睬})}

Ques—questions an entity, e.g. 誰 def:{Ques({human|人})}

2.1.3 The differences between E-HowNet Ontology and HowNet Ontology

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 TopNode. There are two subtrees, {entity|事物} and {relation}, 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} and {object|物體} 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 sub nodes are redistributed to proper position under the subtree of {object|物體}. For instance, the nodes of country names are moved to the sub nodes of {country|國家}. There is no subtree of relation types in HowNet. To establish a taxonomy for semantic roles we constructed a relational hierarchy which includes semantic roles-for-objects and roles-for-events as well as logical functions. The sub nodes of HowNet {Attribute|屬性} were redistributed to their appropriated places under the E-HowNet subtree of the node {property|性質}.

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

We adopt the taxonomic structures of events and objects of HowNet and made some minor adjustments. The event and object subtrees are constructed majorly by is-a (i.e. hypernym-hyponym) relations, and part-whole relation, type-instance relation in some cases to form an inherent system. For example, {BodySubstance|身體部件} under {AnimalHuman|動物} and {孔子|Confucius} under {思想家|philosopher}. Semantic features are attached to events and objects. For instance, the argument structures are attached to different types of events. The subtrees of attribute-values of HowNet, defines the range values of each respective attribute, are classified into {PropertyValue|特性值} and {SituationValue|狀況值} under {state|狀態} of {event|事件}. 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 are all attached to the corresponding objects or events. For example, {KinshipFunction|親屬函數} under {human|人}, {TimePointFunction|時間函數} under {TimePointValue|時間值} and {Direction/Location/PositionValueFunction|方向方位特性值函數} under {LocationalValue|方位值}.

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}, {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 an 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 possibility, necessity and AsExpected 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 (15) respectively shows the usage of agentive, purpose, constitutive and formal:

(11) *premature baby* 早產兒

def:{human|人:age={child|少兒},agentive={labour|臨產:TimeFeature={early|早}}}

(12) *dog food* 狗食

def: {food|食品:telic={feed|餵:target={狗|dog}}}

(13) *wooden stick* 木棍

def: {棍子|stick:material={wood|木}}

(14) *rosy clouds* 彩霞

def: {CloudMist|雲霧:color={colored|彩}}

(15) *spicy and sour soup* 酸辣湯

def: {湯|soup:taste={and({sour|酸},{peppery|辣})}}

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 agentive properties are usually represented by eventive features, which are event frames. For instances , the concepts of *teacher* 老師 and *student* 學生 may be defined and differentiated as *teacher* 老師 def={human|人: telic={teach|教:agent={~}}} and *student* 學生 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:

(16) <request-appeal 請求> def: {commu-reques|求取: manner= {formal|正式}}

(17) <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. Such kind of inheriting property is also held in E-HowNet.

2.2.2 Use basic concepts and relations to define new concepts

HowNet uses a set of primitive semantic units, called sememes, to define concepts. For example, *dog* 狗 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 *Beijing dog* 獅子狗 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 defined by basic or simpler concepts instead of semantic primitives is allowed and all attribute relations are explicitly expressed. The well-defined simpler concepts are called basic concepts which are consisting of a Chinese word ahead and its English equivalent followed. For instance, in E-HowNet *Beijing dog* 獅子狗 is defined as def:{狗|dog:source={北京|Beijing}}. With the basic concept '狗|dog' as the head sense, it denotes the hypernym-hyponym relation between ‘dog’ and ‘Beijing dog’. Hence the definitions of E-HowNet are self-organized as an ontological network.

In order to achieve unambiguous and language-independent definitions, E-HowNet adopts WordNet synsets as an alternative vocabulary for conceptual indexing and representation. Take (18) for example:

(18) *exhibit as evidence* 證物

a. Original E-HowNet definition

def:{inanimate|無生物:
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+corroborate>:
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.

(19) *site of a factory* 廠址

def: {location({工廠|factory})}
def1:{location({InstitutePlace|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 can be defined by any well-defined concepts and 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 *department of literature* 文學系 is like (20):

(20) *department of literature* 文學系
def: {科系|department: predication={and({teach|教},{study|學習}):
location={~}, content={literature|文}}}.

Since the concept {科系|department} is a well-defined basic concept, the above definition can be further extended into the primitive level definition (20'). The notation of '~', as in HowNet, refers to the head concept of the definition which is {科系|department} in (20). Note that the feature of 'predication={and({teach|教},{study|學習}): location = {~}}' in (20') is redundant and will be eliminated after feature unification process (cf. section 4).

(20') def:{PartOf({InstitutePlace|場所:qualification={HighRank|高等},telic={and({teach|教},{study|學習}):location={~},domain={education|教育}}})}.

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* 狗 is not a sememe, but it can be used to describe all sorts of different dogs, such as:

(21) *Great Dane* 大丹狗

def1: {狗|dog: telic={狩獵|hunt:instrument={~},size={big|大},property ={gentle|柔}, color={ and({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}, property={gentle|柔}, color={and({black|黑},{white|白})}}

(22) *mast* 主桅桿

def1: {桅|mast:telic={hang|懸掛:theme={帆|sail},location={~}},qualification={important|重要}}

def2: {PartOf({ship|船}):telic={hang|懸掛:theme={PartOf({ship|船}): telic={drive|駕馭: 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 basic 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, *by* 被 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.

Function words	Content words
Relational senses ←--→	Content senses
de, prepositions, conjunctions, adverbs, ……………………,	adjectives, verbs, nouns

Table 2. The sense spectrum for syntactic categories

Word	POS	Definition
because 因為	Cb	cause ={ }
rain 下雨	Va	{WeatherBad|壞天}

clothes 衣服	Na	{clothing|衣物}
all 都	Da	quantity={all|全}
wet 濕	Vh	{wet|濕}
le 了	Ta	aspect={Vachieve|達成}

Table 3. Examples of E-HowNet lexical sense representations

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 hypernymn 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).

(23) *vehicle headlight* 車燈 def: {PartOf({LandVehicle|車}):telic={illuminate|照射: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.

(24) *everything* 凡事 def: {fact|事情:quantifier={generic|通指}}

(25) *The Nankang area* 南港區 def: {district|區:name={"南港"},quantifier={definite|定指},location={台北|Taipei}}

(26) *neutral zone* 中立區 def: {district|區:predication={中立|neutral:location={~}}}

Words expressed by kinship relations

(27) *father's married sister* 姑媽 def:
{sister({father({human|人})}):qualification={married|已婚}}

(28) *male cousins of different surname* 表兄弟 def: {cousin({human|
人}):gender={male|男}}
def: {or({son({sibling({mother({human|人})})})},
{son({sister({father({human|人})})})})}

Temporal related words

(29) *first love* 初戀 def: {affairs|事務:CoEvent={love|愛戀:sequence={first|首次}}}

(30) *determine victory* 決勝 def: {HaveContest|較量:sequence={last|最後}}

(31) *night sky* 夜空 def: {sky|空域:duration={night|夜}}

(32) *night trip* 夜行 def: {function|活動: duration={night|夜}}

(33) *ancient costume* 古裝 def: {clothing|衣物:telic={PutOn|穿
戴:TimePoint={past|過去},theme={~}}}

(34) *Death may come any minute* 命在旦夕
def: {die|死: TimePoint={TimeAfter({SpeakingTime|說話時間})},
TimeFeature={TimeNear|時間近}}

(35) *person who goes to bed late* 夜貓 def: {human|人:predication={sleep|睡:
agent={~}, TimeFeature={late|遲}}}

(36) *suddenly arrive* 驟至 def: {arrive|到達:manner={sudden|驟然}}

(37) *love develops with time* 日久生情 def: {love|愛戀:cause={associate|交往:
duration={TimeLong|長時間}}}

(38) *happen once in a blue moon* 千載難逢 def: {happen|發生:frequency={rarely|
偶爾}}

(39) *end of month* 月底 def: {month|月:TimeFeature={ending|末}}

(40) *flower season* 花期 def: {TimeSection|時段:predication={開花 |blossom:duration={~}}}

(41) *desert* 始亂終棄 def: {abandon|放棄:manner={wicked|歹}, TimePoint={TimeAfter({love|愛戀})}}

Spatial concepts (Shih et al., 2005)

- Place nouns

Specific place

(42) *Taipei* 台北 def:{capital|國都:location={Taiwan|台灣},quantifier={definite|定指},name={"台北"}}

General place

(43) *a study* 書房 def:{room|房間:telic={study|學習:location={~}}}

(44) *post office* 郵局 def: {institution|機構:telic={post|郵寄:location={~}}}

Specific part of place

(45) *in study room* 書房內 def: {internal({書房|studio})}

(46) *space between fingers* 指縫 def: {縫|chink:position={InBetween({手指})}}

(47) *behind the house* 屋後 def: {hind({house|房屋})}

(48) *the river side* 河畔 def: {side({河|river})}

- Place adverbs

(49) *everywhere* 到處 def: location={WholePlace({object|物體})}

(50) *He hides everywhere* 他到處躲藏 def:{hide|藏匿: agent={3rdPerson|他人}, location={WholePlace({object|物體})}}

(51) *along the street* 沿街 def: LocationThru={route|道路}

(52) *He peddles along the street* 他沿街叫賣 def:{sell|賣:LocationThru={route|道路}, means={cry|喊}}

- Place Prepositions

(53) *from* 從 def: LocationIni={}

(54) *I come from the mountain* 我從山中來 def:{come|來:theme={speaker|說話者}, LocationIni={internal({mountain|山})}}

Shape-related words:

- appearance

(55) *shape of a train* 火車狀 def:appearance={appearance({火車|train})}

(56) *doll face* 娃娃臉 def:appearance={appearance({AnimalHuman|動物:age={young|青年}})}

- shape

(57) *stripe* 條狀 def: shape={cubic|體:length={LengthLong|長}}

(58) *blister* 泡狀 def: shape={cubic|體:shape={round|圓},weight={NotHeavy|輕}}

(59) *globular* 球狀 def: shape={cubic|體:shape={round|圓}}

(60) *filiform* 絲狀 def: shape={linear|線:qualification={fine|纖}}

(61) *lump* 塊狀 def:shape={cubic|體:qualification={GeoIrregular|形狀不規則}}

(62) *droplet-shaped* 滴狀 shape={cubic|體:shape={round|圓},size={small|小}}

(63) *ellipse* 橢圓 def:shape={round|圓:degree={ish|稍}}

(64) *rectangular* 長方 def:shape={square|方:length={LengthLong|長}}

(65) *triangle* 三角 def:shape={angular|角:quantity={3}}

Negation expression

- Negative Polarity→ not

(66) *cross-party* 跨黨 def: {not({distinguish|分辨}):patient={政黨|PoliticalParty}}

- Gradable sense →degree

(67) *short to long length* 中長程 def: distance={ far|遠: degree={ish|稍}}

(68) *brand new* 全新 def: {new|新:degree={extreme|極}}

(69) *I don't really like it.* 我不怎麼喜歡它 def: {FondOf|喜歡:experiencer={speaker|說話者},content={3rdPerson|他人}, degree={ish|稍}}

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

(70) *cut into*切成 def: {become|成為:result={x/{object|物體}},cause={cut|切削:patient={x}}}[2]

(71) *fill up* 加滿 def: {fill|填入:theme={x/{object|物體}},result={full|滿:theme={x}}}

- Subject-control verb-result compounds

(72) *climb out* 爬出來 def: {crawl|爬:direction={external({object|物體})}}

(73) *flow through* 流過 def: {flow|流:LocationThru={}}

(74) *cloy* 吃膩 def: {disgust|厭惡:cause={eat|吃:frequency={often|經常}}}

(75) *conclude something to* 歸納到 def: {classify|分類:component={information|訊息},LocationFin={}}

(76) *merge* 併為 def: {become|成為:means={merge|合併}}

(77) *stand up* 爬起來 def: {arise|起身:result={stand|站立}}

(78) *pull up* 拉上來 def: {pull|拉:direction={upper({object|物體})}}

(79) *jump down* 跳下去 def:{jump|跳:result={GoDown|下去}}

Causative expression

(80) *benign cause* 善因 def: {cause|原因:qualification={positive|正面}}

(81) *pathogen* 病原 def: {cause({ill|病態})}

---

[2] The symbol "x" is a co-index label and the slash "/" denotes semantic restriction. The detail usages of co-index label will be described in section 2.4.

(82) *appetite* 食慾 def: {aspiration|意願:predication={eat|吃:cause={~}}}

Aspectual expression

- Perfective

(83) *finish eating* 吃完 def:{eat|吃:aspect={Vachieve|達成}}

(84) *lift up* 抬起來 def: {lift|提升:aspect={Vachieve|達成}

(85) *formulate* 制定出來 def: {forming|形成:aspect={Vachieve|達成}

(86) *tie up something* 扣好 def: {fasten|拴連:means={press|按壓},aspect={Vachieve|達成}}

(87) *overhaul* 追趕上 def: {chase|追趕:aspect={Vachieve|達成}

(88) *catch* 捉到 def: {catch|捉住:aspect={Vachieve|達成}

- Durative

(89) *listen* 聽下去 def: {listen|聽:aspect={Vgoingon|進展}}

- Experiential

(90) *has read* 讀過 def:{read|讀:aspect={Vachieve|達成}}

- Delimitative

(91) *stare blankly in short time* 呆了呆 def:{stupefied|木然:duration={TimeShort|短時間}}

(92) *try to count* 寫寫看 def:{try|嘗試:content={write|寫}}

Pragmatic expression

(93) *your wife* 尊夫人 def: {wife({listener|聽者}): SpeakerAttitude={respect|敬佩}}

(94) *my wife* 賤內 def: {wife({speaker|說話者}): SpeakerAttitude={modest|謙}}

(95) *talk nonsense* 放屁 def: {TalkNonsense|瞎說: SpeakerAttitude={ExpressAgainst|譴責}}

(96) *wise man won't fight against impossible odds* 好漢不吃眼前虧 def: {surrender|屈服: SpeakerAttitude={persuade|勸說}}

Proper noun

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

(97) *Qixing Mountain* 七星山 def: {山|mountain:quantifier={definite|定指},name={"七星山"},location={Taiwan|台灣}}

(98) *Mukden incident of September 18th, 1931* 九一八事變 def: {事變|incident:quantifier={definite|定指},name={"九一八事變"},location={China|中國}}

Pronoun

(99) *I* 我 def: {speaker|說話者}

(100) *you* 你 def: {listener|聽者}

(101) *he* 他 def: {3rdPerson|他人:gender={male|男}}

(102) *they (female)* 她們 def: {3rdPerson|他人:gender={female|女}, quantity={mass|眾}}

(103) *your father* 乃父 def: {father({listener|聽者}) }

(104) *your honor* 您 def:{listener|聽者: SpeakerAttitude={respect|敬佩}}

(105) *I,this lowly official* 下官 def: {speaker|說話者:apposition={official|官},attitude({~})={modest|謙}}

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 two different types of modalities, i.e. epistemic and deontic. However, some researcher adopted a more open perspective (Hwang 1999, Li 2003, Hsieh 2003, Hsieh 2005) which admits capability, volition and expectation are also within modal categories because in a certain extent they are in line with the feature of "evaluating some piece of knowledge in a possible world" on semantic grounds, but not necessary auxiliaries. In E-HowNet, only epistemic, deontic, and AsExpected senses are regarded as pure modality since lexemes with these three senses are mostly adverbs. The other three modal senses of ability, willingness and expectedness and their verbal counter meanings are represented as attribute or mental verbs, illustrated as Figure 2..

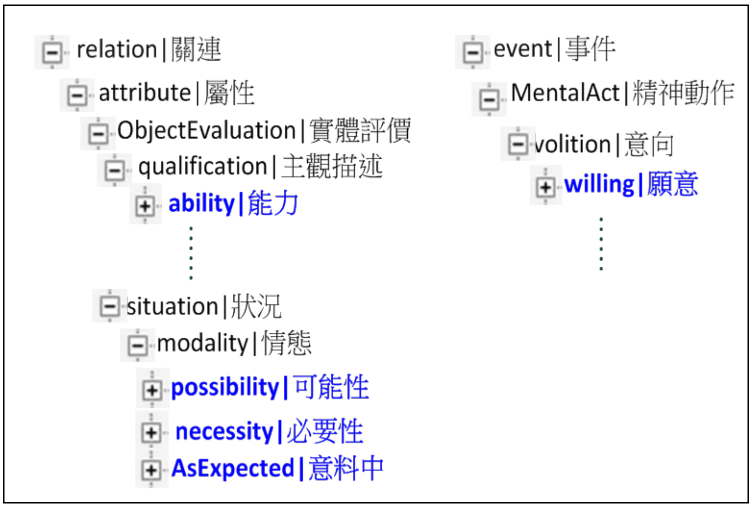

Figure 2: Modal Categories in E-HowNet Sense Representation System

(106) *He is impossible to come.* 他不可能來 def:{come|來:theme= {3rdPerson|他人: gender={male|男}}, possibility={least|無}}

(107) *you don't need to come.* 你不必來 def: {come|來:theme={listener|聽者}, necessity={ish|稍}}

(108) *He is unable to come.* 他沒法參加 def:{參與|ParticipateIn:theme={3rdPerson|他人: gender={male|男}},ability={least|無}}

(109) *He run away unexpectedly.* 他竟然逃走 def:{flee|逃跑:agent={3rdPerson|他人: gender={male|男}},AsExpected={ish|稍}}

(110) *He would rather starve.* 他寧可挨餓 def:{ willing|願意:content={HungryThirsty|飢渴:experiencer={3rdPerson|他人: gender={male|男}}}, degree={very|很}}

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 3):

(111) *because* 因為 def: cause={}

(112) *therefore* 所以 def: result={}

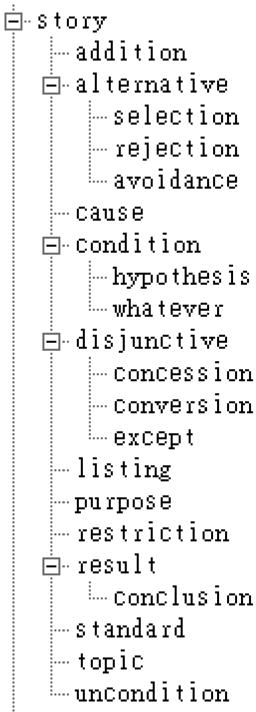

Figure 3. Taxonomy of conjunctions

Adverb

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

(113) *extremely* 透頂 def: degree={very|很}

(114) *Watermelon cold downright* 西瓜冰涼透頂 def:{ chilly|涼:degree={very|很}, theme={西瓜|watermelon }}

Prepositions

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

(115) *from or to* 向: def: source={}, e.g. 台北向印尼購買天然氣；def: target={},e.g. 本黨鄭重向全台灣人民宣布；def: direction={}, e.g. 而車子還是不知道開向何方

(116) *rely on* 藉由: def: instrument={object|物體}, e.g. 藉由此蒸氣孔釋出多餘的能量；def: means={event|事件}, e.g. 希望藉由外派，讓公司幫他辦移民

(117) *after an interval of / at a distance from* 隔: def: TimePoint={TimeAfter()}, e.g. 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={}}, e.g. 秦軍和晉軍隔著肥水遙遙相對

Question words

(118) *why* 為何 def: reason={Ques |疑問}

(119) *how many* 多少 def: quantity={Ques|疑問}

(120) *why* 怎麼 def: cause={Ques|疑問}

Non-predicative Adjective [3]

(121) *bottled* 瓶裝 def: predication={ wrap|包紮:instrument={瓶子 |bottle },patient={~}}

(122) *all-purpose* 萬用 def: qualification={various|多種:theme={intention|意圖}}

(123) *for business use* 商用 def: predication={use|利用:domain={economy|經濟}, instrument={~}}

(124) *medical* 藥用 def: purpose={doctor|醫治}

(125) *continental* 歐式 def: source={Europe|歐洲}

[3] Although non-predicative adjectives are content words, they play modifier roles only, so they are represented by relational sememes.

(126) *patrimonial* 祖傳 def: source={forefathers|祖先}

(127) *cotton made* 棉製 def: material={棉|cotton}

(128) *bacteria-free* 無菌 def: predication={not({exist|存在}):location={~},theme={bacteria|微生物}}

(129) *vegetative* 植物性 def: source={plant|植物}

(130) *in reserve* 備用 def: predication={SetAside|留存:telic={replace|代替}, theme={~}}

(131) *model of palace* 宮殿式 def: predication={alike|似:contrast={皇宮|palace},theme={~}}

(132) *grouped by age* 分齡 def: ComparativeAttribute={age({animate|生物})}

Determinatives

- Demonstrative determinatives

(133) *this* 這 def: quantifier={definite|定指}

(134) *previous* 上 def: sequence={preceding|上次}

- Specifying determinatives

(135) *every* 每 def: range={all|全}

(136) *others* 其他 def: qualification={other|另}

- Numeral determinatives

(137) *one* 一 def: quantity={1}

- Quantitative determinatives

(138) *many* 許多 def: quantity={many|多}

(139) *some* 有的 def: quantity={some|些}

(140) *quite a lot* 之多 def: quantity={approximate()}

Interrogative determinatives (Huang & Chen, 2008)

(141) *what* 啥 def: property={Ques|疑問}

Measure words (Tai et al., 2009)

- Measure words with content sense

(142) *bowl* 碗 def: container={碗| bowl}

(143) *meter* 米 def: length={公尺|m}

(144) *month* 月 def: duration={month|月}

(145) *kind* 樣 def: {kind({object|物體})}

- Measure words without content sense

(146) *copy* 本 def: {null|無義}

(147) *room* 間 def: {null|無義}

(148) *measure word of event* 宗 def: {null|無義}

2.4 Syntax of E-HowNet expressions

The syntax of E-HowNet expressions (sense representations and definitions) 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 include primitive concepts (sememes), basic concepts and complex concepts all expressed by E-HowNet expressions. Relations are semantic roles. Functions are members of Function types in E-HowNet ontology, such as not(), and(,), or(,), PartOf(), Ques(), father(), east(),…etc.. Variables are {~, X, X1, X2, ...,Y, Y1, Y2,…} which are used in E-HowNet expressions for co-indexing entities. "~" denotes (co-indexing with) the highest level head concept of the expression, for example *post office* 郵局 def: {institution|機構:telic={post|郵寄:location={~}}}. The variables X, X1, X2, Y, Y1, Y2 are for co-indexing. The X co-index denotes identical entities. The Y co-index denotes same type. For instances, 母子之情 def: {emotion|情感:predication={associate|交往:agent={relatives(x1/{human|人},x2/{human|人}):mother({x1})={x2}},cause={~}}}. The constants refer to some particular instances, such as a proper name "台北" (e.g. *Taipei* 台北 def: {capital|國都:location={台灣|Taiwan},quantifier={definite|定指},name={"台北"}}) and individual constants, such as {speaker|說話者, listener|聽者, 3rdPerson|他人, SpeakingTime|說話時間…} which are considered as primitive concepts too. The following symbols {‘:’, ‘=’, ‘/’, ‘(‘, ‘)’, ‘{‘,’}’, ‘,’} are delimiters for E-HowNet expressions.

‘:’ is the delimiter between head concept and its following features.

‘=’ denotes the value equivalence.

‘/’ marks a semantic restriction. For instance, *sever* 切斷 def: {cut|切削:patient={x/{object|物體}},result={separate|分離:companion={x}}}

‘(‘, ‘)’ are bracketing symbols for functional arguments.

‘{‘,’}’ are bracketing symbols for a concept.

‘,’ is a delimeter separating two features.

Table 4 shows the basic expressions of E-HowNet.

Concept def:= {Hypernym : Feature,…, Feature}, or {Concept} or {Sememe};	The expression means that a concept may be defined by (a) its hypernymous concept and semantic features, or (b) a synonymous concept, or (c) a primitive concept.
Features def:= Relation(x)={Concept};	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 concepts and variables. Relation(~)={Concept} will be abbreviated as Relation= {Concept}.
Relation def:= property, content, host, location, agent, patient,….;	a set of semantic relations.

Table 4. 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.

(149) *Great Dane* 大丹狗

def:{狗|dog:

location={German|德國},

telic={hunt|狩獵:instrument={~}},

size={big|大型},

property={gentle|溫和},

color={and({black|黑},{white|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:

(150) *hunting gun* 獵槍

def:{gun|槍:

telic={hunt|狩獵:

instrument={~}}}

(151) *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={or({black|黑},{white|白})}} which happens to be the features shared by the following examples:

(152) *panda* 貓熊

def:{beast|走獸:

location={China|中國},

predication={eat|吃:

patient={竹|bamboo},

agent={~}},

color={and({black|黑},{white|白})}}

(153) *dairy cow* 乳牛

def: {牛|cattle:

telic={take|取:

theme={奶|milk},

source={~}},

color={and({black|黑},{white|白})}}

c. Achieves near canonical semantic representation.

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

(154) 我 買了 一本 科幻小說。

(155) *I bought a science fiction.*

Both sentences have the same representation of {buy|買: agent={speaker|說話者}, possession={小說|fiction: qualification={or({scientific|科學},{fake|偽})},quantity={1}}, TimePoint={TimeBefore({SpeakingTime|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.

(156) *tailor store* 裁縫店 def: {商店|store: telic={裁縫|sew: location={~}}} can be

extended to:

{InstitutePlace|場所:

telic={produce|製造:

PatientProduct={clothing|衣物},

means={or({剪|ToCut},{fasten|拴連})}},

location={~}},

domain={economy|經濟}}

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 {店| store}.

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:

(157) *bulletin board* 公佈欄

def1: {facilities|設施:telic={announce|發表:location={~}}}

def2:{[(establishment)]:telic={ [(announce, denote)]:location={~}}}

f. Rather than creating a completely new ontology, E-HowNet accommodates existing ontologies like WordNet, HowNet, and FrameNet.

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, example (20') showing the decomposition result of (20), the hypernymous class {科系| department} of {文學系} is not a primitive concept and is decomposed into the definition of {InstitutePlace|場所:domain={education|教育},predication={and({teach|教},{study|學習}):location={~}}}. Then, the reduplicated features of predication={and({teach|教},{study|學習}):location={~}} appear after the decomposition process, as in (20'). Finally, the reduplicated features will be unified into a single feature of 'predication={and({teach|教},{study|學習}):location={~}, content= {literature|文}}' and the ground level representation of {文學系}becomes:

(20')　def:{InstitutePlace|場所:

domain={education|教育},

predication={and({teach|教},{study|學習}):

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 example, we will show how step 1 and 2 are done:

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

In (158), *wet* 濕, *clothes* 衣服 and *rain* 下雨 are content words, whereas *all* 都, *le* 了 and *because* 因為 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 (159) are derived by parsing sentence (158):

(159) S(cause:VP(Head:Cb:因為|Dummy:VA:下雨)|theme:NP(Head:Na:衣服) | quantity: Da:都 | Head:VH:濕|particle:Ta:了)。

Then, (160) is the semantic composition which results from the unification process. The dependency daughters become feature attributes of the sentential head *wet* 濕.

(160) def:{wet|濕:
theme={clothing|衣物},
aspect={Vachieve|達成},
quantity={all|全},
cause={rain|下雨}}.

In (160), the function word *because* 因為 links the head concept *wet* 濕 and *rain* 下雨 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.

From the above discussion, it revealed that the basic semantic composition process is combining syntactic dependent constituents step by step and produces a series of compositional semantic expressions of triples. Where a triple is a basic compositional semantic unit composed by host, attribute, and value and expressed as attribute(host)={value}. For instances, color({lights|光})={blue|藍}, agent({publish|出版})={ human|人}. Generally host role is played by entities and attributes are so-called relations. Therefore E-HowNet ontology consists of two major subtrees. One is entity subtree and another is relation subtree. The entity tree contains all concepts which might play the role of host. Note that since most of relations may also play the role of host, relation-entities are also duplicated in the entity tree.

Some attributes may have specific range of values. For instance, values of color are red, blue, and yellow etc. In E-HowNet, attributes and their respective values are constructed in parallel. Such information is very useful in identifying semantic relations between two constituents while doing semantic composition. To facilitate automatic semantic composition and language understanding, E-HowNet ontology provides additional lexical information other than conceptual definitions and part-of-speeches. The additional lexical information is described below and Figure 4 shows the lexical information of 海量.

詞彙訊息

<table>
<tr><td>詞彙:</td><td>海量</td></tr>
<tr><td>詞性:</td><td>Nad</td></tr>
<tr><td>英文意涵:</td><td>great capacity for liquor</td></tr>
<tr><td>Definition: [編輯]</td><td>{ability({飲酒|DrinkLiquor}):value={extreme|極}}</td></tr>
<tr><td>操作式:</td><td>ability({飲酒|DrinkLiquor})={extreme|極}</td></tr>
<tr><td>語義功能:</td><td>AbilityValue|能力值</td></tr>
<tr><td>展開式:</td><td></td></tr>
<tr><td>WordNet 自動連結:</td><td>{great.n.01, liquor.n.01, liquor.n.03, liquor.n.02, capacitance.n.01}</td></tr>
</table>

Figure 4. The lexical information of “海量”

4.1 Other lexical information to facilitate semantic composition and language understanding

a) Syntactic and semantic functions

As we had mentioned a lexical word may play different syntactic and semantic functions, it may ambiguously denote many lexical concepts. In E-HowNet each lexical concept of a word is identified and provided with its sense definition, English translation, part-of-speech, and major semantic functions. Since all entities may play the semantic function of host, only semantic function roles of attribute and value will be provided. For instance, 貪腐 Pos: VH11; Translation: embezzle and corrupt; Def: {and({acquisitive|貪心},{immoral|不道德})}; Semantic function:

AcquisitivenessValue|貪慾值.

b) Additional lexical sense expressions for compositional processing

A lexical concept (word) may also play different grammatical functions. For instance, a stative verb may play the role of subject/object, predicate, modifier etc. Other than basic semantic expression, we like to know its event frame (i.e. arguments) while it plays the predicate role. On the other hand while playing modifier role, we need to know what the relation between modifier and head is. Therefore in E-HowNet, for the lexical concepts playing multiple roles, additional sense expressions may be provided to facilitate semantic composition processing. For instance, 役齡 def: {age|年齡:qualification={當兵|ServeInArmy}} and an additional definition for operation is also provided def: age={age|年齡:qualification={當兵|ServeInArmy}} which is in the form of playing the role of modifier.

c) Event frames

Arguments of each event type are provided.
e.g. infect|傳染
ACTOR{causer},THEME:*disease passes on to GOAL*{theme},GOAL{target}
e.g. install|安裝
ACTOR:*installer*{agent},THEME:*thing installed*{theme},LOCATION{location}

d) Semantic links

E-HowNet ontology is constructed by is-a relation which has the inherent property. Hyponym concepts inherit the properties of hypernym concepts. There are also many other important relations other than is-a relation among concepts. Chapter 9 of [Dong & Dong 2006] states that there are 11 types of explicit relations in HowNet. They are the relations of synonym, synclass, antonym, converse, hypernym, hyponym, part-to-whole, value-to-attribute, attribute-to-host, cognate role-frame, semantic-roles-to-event. Those semantic links are also maintained in the E-HowNet. Below is the information of the primitive {include|納入} containing sematic links with other event primitves.

義原訊息

Primitive:	include|納入
Definition:	{AlterInclusion|變包含:result={contain|包含}}
Event frame	ACTOR{agent},THEME:*entity put into*{component},LOCATION:*location put into*{whole}
Primitive relations:	converse={discharge|開除}; converse={withdraw|退出}; implication={BeMember|隸屬}; implication={contain|包含}
Conflation of events:	BeMember -> component({include|納入})=theme({BeMember|隸屬}); BeMember -> whole({include|納入})=source({BeMember|隸屬}); contain -> component({include|納入})=content({contain|包含}); contain -> whole({include|納入})=theme({contain|包含})
WordNet 連結:	{include} (同義)
WordNet 自動連結:	{entity.n.01, abstraction.n.06, have.v.01, keep.v.03, record.v.01}

5. Potential Applications of E-HowNet

There is still a long way to go in order to achieve fully automatic semantic composition and natural language understanding. Many research problems and difficulties need to be solved, 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 indispensable 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 *hired herdsman* 牧工, *nuclear industry* 核工,or *art of singing* 唱工, 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 5.

The meaning of *herdsman* 牧工, *nuclear industry* 核工 ,or *art of singing* 唱工 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 *nuclear industry* 核工 by replacing the value of feature 'domain' with the sense of '*nuclear* 核' to create a new definition as (161):

(161) *nuclear industry* 核工

def:{industry|工業:domain={nucleonics|核子學}}

In similar way, *art of singing* 唱工 can be defined as (162):

(162) *art of sewing* 縫工

def: ability({縫紉|sew})

Sense	example	E-HowNet definition
labor 工人	*porter* 搬運工	def:{labor|工人:telic={transport|運送: theme={goods|貨物},agent={~}}}
	female labor 女工	def:{labor|工人:gender={female|女}}
	child labor 童工	def:{labor|工人:age={child|幼兒}}
industry 工業	*chemical industry* 化工	def:{industry|工業: domain={chemistry|化學}}
	engineering industry 機工	def: {industry|工業:domain={machine|機器}}
skill 技術	*cutting skill* 刀工	def: ability({cut|切削})
	painting skill 畫工	def: ability({draw|畫})

Table 5. The senses of morpheme "工" and examples for each sense

5.2 Sense disambiguation

E-HowNet's lexical sense definitions provide many binary conceptual relations which are sources of world knowledge and can be utilized for sense disambiguation. In addition, semantic restrictions for the values of the relations marked by function words were also encoded. For instance, the preposition *from* 從 has two different relational senses and each denotes the relation below:

(163) *from* 從

def: location-LocationIni={place|地方}

def: TimePoint-TimeIni={time|時間}

The sense of *from* 從 can be disambiguated by the respective semantic restrictions either {place|地方} or {time|時間} 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 *fast* 快 and *food* 食品 between *fast* 快 and *car* 車 are different and cannot simply be described as property-entity relation. The semantic gaps regarding *serving fast* and *moving fast* respect to two compounds are not expressed explicitly. The different telic feature values for *food* 食品 and *car* 車 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 (164)~(166), we demonstrate how to restore sense omissions by object-attribute relations.

(164) *I like the red (something)* 我喜歡紅的

def:{FondOf|喜歡:
experiencer={speaker|說話者},
target={object|物體:
color={red|紅}}}.

Because the semantic role ‘color’ is an attribute of objects, it implies an object was missing in the sentence (164) 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 (165) thus signifies the presence of an object. For the same reason as exemplified in (164), we know the object is omitted in (166) too. The event frame of {speak|說} has been coded as to take {human|人} as an agent role. We therefore know the absent object has to be an instance of {human|人}.

(165) *few* 少數
def: quantity={few|少}.

(166) *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 *bi* 比. The sense of *bi* comprises a complex argument structure which is shown in (167). Sentence (168) is its implementation:

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

(168) *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={3rdPerson|他人},
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. (169) shows the semantic representation with the semantic gap filled in:

Fine-grained Semantic Roles	**Thematic Roles**	**Grammatical Functions**
Profiled_Item+(Profiled_Attribute)	Theme; Experiencer	Subject
Standard_Item+(Standard_Attribute) Comparison_set	Contrast	Object[PP[bi]]
Attribute_Value	Head	Verb
Degree	ComparativeQuantity; Degree	Complement
Manner	Manner	Adjunct (Manner)
Place	Location	Adjunct (Location)
Time	Time	Adjunct (Time)

Table 6. Mapping table for the fine-grained semantic roles

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

def:{tall|高:

Profiled_Item={speaker|說話者},

Profiled_Attribute={height|高度},

Standard_Item={3rdPerson|他人},

Standard_Attribute={height|高度},

Degree={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.

(170) 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:

def:{抓獲:agent={機長},patient={搶犯:gender={女}},manner={機敏}}

vs.

def:{逮捕:agent={飛機正駕駛},patient={強盜:gender={女}},manner={敏捷}}

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

def: {catch|捉住:agent={official|官: 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={clever|靈}}

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|買} and {sell|賣} 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, such that meaning representations for morphemes, words, phrases, and sentences can be uniformly represented under the same framework. New concepts can be defined by previously known concepts and definitions can be dynamically decomposed into lower level representations until the ground-level definition is reached. Near-canonical representation thus can be achieved at a suitable level of representation for synonyms or paraphrases. 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 misunderstandings, 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

Baker, C. F., C.J. Fillmore, & J.B. Lowe, 1998,"The Berkeley FrameNet Project", *Proceedings of Coling 1998*, pp. 86-90.

Chen, Keh-Jiann & Chao-Jan Chen, 2000,"Automatic Semantic Classification for Chinese Unknown Compound Nouns," *Proceedings of Coling 2000*, pp.173-179.

Chen, Keh-Jian & Ming-Hong Bai, 1998, "Unknown Word Detection for Chinese by a Corpus-based Learning Method," *International Journal of Computational linguistics and Chinese Language Processing*, vol.3(1), pp.27-44.

Chen, Keh-Jiann, Shu-Ling Huang & Yueh-Yin Shih, Yi-Jun Chen, 2005, "Extended-HowNet- A Representational Framework for Concepts," OntoLex 2005 - Ontologies and Lexical Resources IJCNLP-05 Workshop, Jeju Island, South Korea

Chen, Keh-Jiann, Shu-Ling Huang, Yueh-Yin Shih & Yi-Jun Chen, 2004, "Multi-level Definitions and Complex Relations in Extended-HowNet," Workshop on Chinese Lexical Semantics, Beijing University. (in Chinese)

Chen, Yi-Jun, Shu-Ling Huang, Yueh-Yin Shih & Keh-Jiann Chen, 2005, "Semantic Representation and Definitions for Function Words in Extended-HowNet," Workshop on Chinese Lexical Semantics, Xiamen University. (in Chinese)

Chung, You-Shan, Shu-Ling Huang & Keh-Jiann Chen, 2007, "Modality and Modal Sense Representation in E-HowNet," *Proceedings of Paclic 21*, pp.136-145.

Dong, Zhendong & Qiang Dong, 2006, *HowNet and the Computation of Meaning*, World Scientific Publishing Co. Pte. Ltd.

Dong, Zhengdong & Dong Qiang, HowNet, http://www.keenage.com/

Fellbaum, Christiane, 1998, *WORDNET-An Electronic Lexical Database*, the MIT Press, Cambridge.

Fillmore, Charles, *FrameNet*, http://www.icsi.berkeley.edu/~framenet/

Huang, Shu-Ling & Keh-Jiann Chen, 2008, “Knowledge Representation and Sense Disambiguation for Interrogatives in E-HowNet,“ *International Journal of Computational Linguistics & Chinese language Processing* (IJCLCLP), vol. 13(3), pp. 255-278.

Huang, Shu-Ling & Keh-Jiann Chen, 2009, “Knowledge Representation for Time Intervals in E-HowNet,” CLSW 2009, Yantai, China.

Huang, Shu-Ling, Shih Yueh-Yin & Keh-Jiann Chen, 2008, “Knowledge Representation for Comparison Words in Extended-HowNet,” *Language and Linguistics*, vol. 9(2), pp. 395-414.

Huang, Shu-Ling, You-Shan Chung & Keh-Jiann Chen, 2008, “E-HowNet: the Expansion of HowNet, the First National HowNet Workshop,” Beijing, China.

Huang, Shu-Ling, Yu-Ming Hsieh, Su-Chu Lin, and Keh-Jiann Chen. 2014. Resolving the Representational Problems of Polarity and Interaction between Process and State Verbs. Computational Linguistics and Chinese Language Processing (IJCLCLP), Vol. 19, No. 2, pp. 33-52.

Pustejovsky, James, 1995, *The Generative Lexicon*, MIT Press, Cambridge.

Schank, R., 1975, *Conceptual Information Processing*, Elsevier Science, New York.

Shih, Yueh-Yin, Shu-Ling Huang & Keh-Jiann Chen, 2006, “Semantic Representation and Composition for Unknown Compounds in E-HowNet,” *Proceedings of PACLIC 20*, pp.378-382.

Shih, Yueh-Yin, Shu-Ling Huang, Yi-Jun Chen & Keh-Jiann Chen, 2005, “Semantic Representation and Composition for Spatial Concepts in Extended-HowNet,” IEEE International Conference on Natural Language Processing and Knowledge Engineering, Wuhan, China

Tai, Chia-Hung, Jia-Zen Fan, Shu-Ling, Huang & Keh-Jiann Chen, 2009. “Automatic Sense Derivation for Determinative-Measure Compounds under the Framework of E-HowNet,” *International Journal of Computational Linguistics & Chinese Language Processing*, vol. 14(1), pp.19-44..

Veale, T., 2005, ”Analogy Generation with HowNet,” *Proceedings of International*

Joint Conference on Artificial Intelligence, pp.1148-1153.

Wierzbicka, A., 1972, *Semantic Primitives*, Athenaum, Frankfurt.

You, Jia-Ming & Keh-Jiann Chen, 2004, “Automatic Semantic Role Assignment for a Tree Structure, “ *Proceedings of SIGHAN Workshop*.

劉群 & 李素建，“基於《知網》的辭彙語義相似度計算,” *International Journal of Computational Linguistics and Chinese Language Processing*, 7(2), pp. 59-76.

Appendix A. The usages of semantic roles

There are three major types of semantic roles, i.e. a) roles for entity, b) roles for object, and c) roles for event. Since objects and events are also entities, the roles for entity are roles for both object and event. Following are examples for the usages of the semantic roles. However due to limited space, some of the sub-nodes of *property, HumanPropensity* and *qualification* are omitted. For the complete taxonomy, please refer to http://ckip.iis.sinica.edu.tw/taxonomy/.

a) Semantic roles for entity

situation—describing situations of entities

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}, ComparativeQuantity={head|頭:quantity={1}}, ComparativeAttribute={stature({human|人})}}

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

frequency —the frequency an event happens, e.g. 反覆思量 def: {think|思考:frequency={again|再}}

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

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. 口授 def: {teach|教:means={speak|說}}

method—the method how an act is done. Most of the time this relation is expressed by the above node "means" and therefore no example of relation in lexicon is found.

price—the price one pays to purchase something, e.g. 生活費 def: {expenditure|費用:telic={alive|活著:price={~}}}

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

sequence—the sequence of object, e.g. 二月份 {month|月:sequence={2}}

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

StateIni—the initial state before an event happens, e.g. 解凍 def: {StateChange|態變:StateIni={ice|冰}}

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}}

locational—the locality of an entity.

location—the location where an object exists or an event happens, e.g. 露營 def: {reside|住下:location={land|陸地:qualification={desolate|荒}},while={tour|旅遊}};
水鳥 def:{bird|禽: location={waters|水域}}

LocationIni—the starting point of a trajectory over which an event takes place,
e.g. 隕石 def: {stone|土石:predication={come|來:LocationIni={celestial|天體},theme={~}}};
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}};
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.來時路 def: {route|道路:predication={arrive|到達:LocationThru={~}}};
流過 def: {flow|流:LocationThru={}}

distance—the distance over which an act takes as its scope or the distance

between an object and a reference point
e.g. 向陽 def: {facing|朝向:direction={太陽|TheSun}};
九霄 def: {sky|空域:height={high|高},distance={far|遠}}

direction—the direction a movement takes or the direction to which an object faces,
e.g. 拉上來 def: {pull|拉: direction={upper({object|物體})}};
放眼望去 def: {look|看:distance={far|遠}},

position—the position where an object occupies,
e.g. 內院 def: {院子|courtyard:position={internal({building|建築物})}}

source—the source of an entity,
e.g. 內憂 def: {mishap|劫難:source={place|地方:source={native|本土}}};
遮陽 def: {block|攔住:theme={lights|光:source={太陽|TheSun}}}

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

TimePoint—a specific time when an event happens or when an object belongs, e.g. 換發 def:{issue|分發: TimePoint={TimeAfter({check|查})};
山頂洞人 def: {human|人:TimePoint={遠古|AncientTimes}}

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. a period of time for which a situation lasts, e.g. 太空時代 def: {dynasty|朝代:predication={LeaveFor|前往:LocationFin={sky|空域},duration={~}}}

TimeFeature—the status of a time point in relation to a reference time, e.g.
期中 def: {學期|SchoolTerm:TimeFeature={middle|間}};
月底 def: {month|月:TimeFeature={ending|末}}

while—an object or an event that exists or occurs at the same time as the main event taking place, e.g. 外出服 def: {clothing|衣物:telic={PutOn|穿戴:while={WhileAway|消閒},theme={~}}};
相視而笑 def: {laugh|笑:while={look|看:manner={EachOther|相互}}

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

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

host — host of attributes, e.g. 地心引力 def: {strength({attract|吸引}):host={earth|大地}}

value — value of attributes, e.g. 台籍 def: {nationality({human|人}):value={台灣|Taiwan}}

b) Semantic roles for objects

possessor—the owner in a possessor-possession relation

member—the member of an object, e.g. 女校 def: {學校|school:member={女性|woman}}

creator—the creator of an object, e.g. 自序 def: {序文|preface:creator={self({作家|writer})}}

owner—the owner of an object, e.g. 官房 def: {宿舍|dormitory:owner={政府|government}}

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={SetAside|留存:theme={鬍鬚|beard:quantity={many|多}}}}

telic—purpose and function of an object, e.g. 甘蔗田 def: {田|farmland:telic={planting|栽植:patient={甘蔗|sugarcane},LocationFin={~}}}

agentive—factors involved in the origin or "bringing about" of an object, e.g. 丹藥 {medicine|藥物:agentive={提煉|ToAbstract:PatientProduct={~},agent={道家|DaoistSchool}}}

property— the property of an object.

age—the age of a living thing, e.g. 雛鳥 def: {bird|禽:age={child|少兒}}

appearance—appearances of an object, e.g. 人形 def: shape={shape({human|人})}

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

weight —the weight of an object, e.g. 重物 def: {inanimate|無生物:weight={heavy|重}}

length—the length of an object, e.g. 半統襪 def: {襪子|socks:length={LengthLong|長}}

kind— object types, e.g. 大千世界 def: {world|世界:kind={various|多種}}

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

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: {route|道路:width={narrow|窄}}

size—size of an object, e.g. 血盆大口 def: {mouth|口: size={big|大}}

taste—taste of an object, e.g. 苦酒 def: {酒|liquor:taste={bitter|苦}}

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

gender—the gender of a living thing, e.g. 女人 def: {human|人:gender={female|女}}

temperature—temperature of an object, e.g. 熱食 def: {edible|食物:temperature={hot|熱}}

dimension—dimensionality of an object, e.g. 次元{dimension({object|物體})}

shape—shape of an object, e.g. 方盒 def: {盒子|box:shape={square|方}}

CoEvent—introduces the event type of an event noun, e.g. 外遇 def: {affairs|事

務:CoEvent={love|愛戀},qualification={immoral|不道德}}

ObjectEvaluation— evaluations of objects.

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

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

PartOf—the relation between an object and its parts or constituents

content— the content of an object, e.g. 二十四孝 def: {故事|story: content={emotion|情感:CoEvent={loyal|忠孝}} }

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

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

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

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})}

BasePart—the base part of an object, e.g. 山麓 def: {BasePart({山})}

EndPart—the end part of an object, e.g. 末梢 def: {EndPart({physical|物質})}

surface—the surface of an object, e.g. 海面 def: {surface({海})}

BodyPart—the body part of an object, e.g. 車體 def: {BodyPart({LandVehicle|車})}

skeleton—the skeleton of an object, e.g.叉枝 def: {skeleton({tree|樹})}

grip—the grip of an object, e.g. 車把 def: {grip({LandVehicle|車})}

passage—the passage of an object, e.g. 走廊 def: {passage({building|建築物})}

edge—the edge of an object, e.g.外緣 def: {edge({inanimate|無生物})}

hole—the hole of an object, e.g. 山洞 def: {hole({山})}

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}}

rate —a specific kind of ratio, in which two measurements are related to each other, e.g. 出生率 def: rate({ComeToWorld|問世})

amount — an enumerable quantity, e.g. 三個 def: amount={3}

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

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

apposition—denoting the equvilent entity or setting examples in context, e.g. 譬如 def: apposition={}

c) Semantic roles for events

story— details of an event.

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

alternative —an alternative way, e.g. 要麼 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={}

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

condition —the condition under which an event happens, e.g. 不請自來 def: {come|來:condition={not({invite|邀請})}}

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

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

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

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={}

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

listing—listings markers in context, e.g.一方面 def: listing={}

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

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

result—the result caused by an event, e.g. 申請 def: {request|要求:result={obtain|得到}}

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

standard— standard of rules, e.g. 太陽曆 def: {law|律法:content={time|時間},standard={太陽|TheSun}}

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

participant— participant roles of an event.

actor—the actor of an event.

agent—a conscious actor which performs an action with control (on purpose) and has a physical, visible effect on object, e.g. 工讀生def: {學生|student:predication={打工|WorkPartTime:agent={~}}}

experiencer—an animate being who perceives a stimulus or registers a particular mental or emotional process or state, e.g. 好戰份子def: {human|人:predication={FondOf|喜歡:target={fight|爭鬥},experiencer={~}}}

causer—an unconscious force which incurs an event without purpose, e.g. 病媒蚊 def: {蚊子|mosquito:telic={infect|傳染:theme={disease|疾病},causer={~}}}

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.流淚 def: {flow|流:theme={BodyFluid|體液: whole={animate|生物},source={eye|眼}}}

product—products of an event.

PatientProduct—an object which comes into physical being after an event, e.g. 製糖 def: {produce|製造:PatientProduct={糖|sugar}}

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={商品|commodity}}

goal—an object which is affected or perceived.

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

patient—the object which is affected, e.g. 殃及池魚 def: {damage|損害:patient={human|人:qualification={irrelevant|不相關}}}

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

source—the source from which a possession is obtained, e.g. 井鹽 def:{鹽|salt:source={井|AWell}}

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

companion—an object which is correspondent to the main body in the events of ‘connective,’ ‘AlterConnection, ’ ‘TimeOrSpace,’ or ‘HaveContest’ types etc, e.g. 上網 def: {connect|連接:companion={internet|因特網}}

contrast—an entity that corresponds to another entity in some way e.g. 過猶不及 def: {equal|相等:contrast={insufficient|不足},theme={more|較}}}

pragmatic—functions of a pragmatic use

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

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

Modality—speaker’s evaluation of an event (Huang, Lin & Chen, 2014)

possibility—the epistemic guessing towards a possible event, e.g. 勝券在握 def: {win|獲勝:possibility={extreme|極}}

necessity—the deontic demand towards a future event, e.g. 他必須來 def:{come|來: agent={3rdperson|他}, necessity={extreme|極}}

AsExpected—the degree of the result achieved as the same as speaker’s expected result, e.g. 果然 def: AsExpected={extreme|極}

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

Appendix B. Formal syntax of E-HowNet Expressions:

Concept → Complex-Concept| Basic-Concept

Complex-Concept → '{' (Basic-Concept| Co-idexed-Concept) ':' Feature-Values '}'|

'{' (Basic-Concept| Co-idexed-Concept) '}'

Basic-Concept→ Sememe| Mapped-Concept |Intermediate-Form| Variable| Constant

Co-indexed-Concept→ Basic-Concept '=' Variable

Variable → '~'| 'X' |'Y' | 'X1'| 'X2' |'Y1' | 'Y2'

Constant → CC* ;

CC→Chinese-Character;

Mapped-Concept → Relation '(' Concept ')'| Function '(' Mapped-Concept')';

Concepts → Concept| Concepts ',' Concept| Null;

Intermediate-Form→ English-Word '|' Chinese-Word | Chinese-Word;

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://ehownet.iis.sinica.edu.tw/index.php.