Concentric: Studies in Linguistics
35.1 (January 2009):87-109

Formal Description of Lexical Semantic Relations

Shu-Kai Hsieh
National Taiwan Normal University

Lexical semantic relations have played an important role in the recent developments of Natural Language Processing and Computational Lexical Resources as well. This paper reviews the notion of lexical semantic relations in the WordNet-like lexical resources, and proposes a formal modeling of lexical semantic relations using the extended Formal Concept Analysis. I believe that the proposed formalization will be able to highlight problems with regard to lexical and cultural gaps, and serve as a foundation for solutions that support lexical theoretical explorations and applications for multilingual wordnets in the future.

Key words: lexical semantics, computational lexicon, Formal Concept Analysis

1. Introduction

In recent years, computational lexical semantics has been recognized as a crucial research field for many Natural Language Processing (NLP) tasks, such as word sense disambiguation, semantic roles labeling, and ontology learning. In particular, there has been an increasing focus on lexicon-driven methodology. Research programs, whose goal is the construction of large-scale lexicons, are all facing with the issues of the appropriate representation structure for different facets of lexical semantic information.

Baldwin (2006) reported three main approaches to lexical semantic categorization: attributional semantic categorization, semantic clustering, and relational semantic categorization. Among these main approaches to lexical semantic categorization, the relational approach – to capture correspondences between lexical items by way of a finite set of pre-defined semantic relations – is probably the least explored. However, in the past few years, lexical knowledge base which systematizes lexical and conceptual information of human knowledge has become a generic language resource for Natural Language Processing applications. Among the emergent language resources, wordnets, pioneered by the Princeton WordNet (Fellbaum 1998), and greatly enriched by EuroWordNet (Vossen 1998), have thus become the de-facto standard for a lexical knowledge base enriched with lexical semantic relations. In addition to the multilingual architecture design of EuroWordNet, there are different computational proposals to construct the expansion for monolingual wordnets to parallel wordnet systems, such as Pianta and Girardi (2002). However, the construction of multilingual wordnets eventually faces the challenge of so-called low-density languages. Low-density languages, as opposed to high-density languages, usually refer to languages which do not have enough existing resources for semi-automatic construction of monolingual wordnets.

By presuming that the essence of lexical semantic relations (LSRs) are more universal than word sense in human languages, previous works (Huang et al. 2002, 2003, 2005) proposed a parallel wordnet bootstrapping strategy based on one monolingual wordnet and a set of presumed cross-lingual LSRs rules. Following the line of thought, we (Hsieh et al. 2006) performed a first large-scale experiment of LSRs acquisition, which has yielded promising results. In order to make the model implementable by all the other low density languages, a formal and solid foundation is urgently needed. This motivates our current study.

In what follows, Section 2 gives a brief summary of lexical semantic relations from different perspectives. Section 3 explains the relations in computational lexical knowledge database, which are the main focus of this study. In Section 4, I propose a notion of LSR algebra along with the extended Formal Concept Analysis approach to the mathematical modeling of lexical semantic relations. I conclude this paper in Section 5.

2. Lexical semantic relations: An overview

Lexical semantic relations (LSRs) have been explored in many disciplines, such as linguistics, anthropology, cognitive science, database design and artificial intelligence. This section presents a brief summary of lexical semantic relations from different perspectives.

2.1 Classification of semantic relations

If a semiotic stance is taken in thinking lexical semantics, meaning will arise from the differences between signifiers. Saussure ([1916] 1983) emphasized that these differences were of two kinds: *paradigmatic* (concerning substitution) and *syntagmatic* (concerning positioning). Paradigmatic relations hold between words of which the meaning is related in some systematic ways. Often they belong to the same syntactic category, as e.g. [*note, short letter, line, billet*]. In contrast, syntagmatic relations are based on the co-occurrence of words within a sentence, like collocations.

From the logical point of view, various types of semantic relations which hold between lexical units can be found. There are four basic types of semantic relations: 1. Identity: LU_1 = LU_2, 2. Inclusion: LU_2 is included into LU_1, 3. Overlap: LU_1 and LU_2

have a non-empty intersection, but either one is not included in the other, and 4. Disjunction: LU_1 and LU_2 have no common element.[1]

From the lexical configuration point of view (Dahlberg 1994, DIN32705 1987), semantic relations can be classified into hierarchical and non-hierarchical relations. Two major types of hierarchical relations are conceptual orderings (also called IS-A relation, taxonomy, class inclusion, hyponymy, or super-ordination), and meronomies (also called part-whole relation).[2] Some other relations that evolve from hierarchies are coordination (relations between co-hyponyms which share the same immediate superordinate), attribution (sometimes called has-a relation).

Among non-hierarchical relations, two functional relations are mainly distinguished, synonymy (overlap in semantic content), and the different forms of oppositions (such as antonymy). Functional relations, which are analogous with syntagmatic relations, form network-like orderings, such as the semantic case relations (agent, instrument, patient) of a verb within a sentence, and the Entity-Relationship-Model of database theory. Oppositions, which are binary sequences, form linear orderings (non-hierarchical).

2.2 Lexical relations and ontological relations

Another important discussion with regard to LSRs studies centers around the lexicalization of LSRs. Previous works of lexical semantics like Cruse (1986) tried to define LSRs by certain ‘linguistic tests’, which were meant to provide the lexical evidence of LSRs. Following this linguistic approach, the EuroWordNet project (Vossen 1998) has proposed a complete list of testing procedures in determining LSRs between lexical units, and has successfully implemented in the resulting multilingual wordnets. The following is an example taken from the project document (Climent et al. 2006).

---

[1] These are called *congruence relations* in Cruse (1986).

[2] In some cases, these two relations are not easy to distinguish. For example, should “computational linguistics” be called a part of linguistics or a kind of linguistics? However, the philosophical debates on these issues are not the focus of this paper.

(1) An example from a project document (Climent et al. 2006)

Test

Comment	General meronymy for nouns
Score	Test sentence
Yes	a (a/an) X makes up a part of (a/an) Y (a/an) Y has (a/an) Xs
No	b the converse of the a) relations
Conditions	X and Y are concrete nouns and are interpreted generically
Effect	X HAS_HOLONYM Y Y HAS_MERONYM X

On the contrary, in the current studies of formal ontologies,[3] ontological relations that have been investigated, such as cause-effect relations, product-producer relations, co-existence relations, temporal and spatial relations, qualia structure relations in *Generative Lexicon* theory (Pustejovsky 1995), focus more on their logical features at the conceptual rather than the lexical level.

In this paper, I focus only on the relations realized in the lexical knowledge base.

3. Relations in the lexical knowledge base

A **lexical knowledge base** is a general repository of knowledge about lexicalized concepts and their relationships. It contains lexical information extracted from machine-readable dictionaries, corpus, data manually obtained from humans. Lexical semantic relations constitute the central element in the organization of lexical semantics knowledge bases. In particular, they have been mainly used to structure the lexicon in the hierarchical organization. They have been extensively used and evaluated in the computational lexical knowledge base, such as WordNet and EuroWordNet.

3.1 WordNet

Now I first look more closely at WordNet, an on-line lexical knowledge base for English developed at Princeton University.[4] Currently, WordNet contains a large set of noun, verbs, adjectives, and adverb synonyms, each representing a lexicalized concept. Words with synonymous senses are allocated in a so-called **synset** (*synonym*

---

[3] For example, The Suggested Upper Merged Ontology (SUMO), http://www.ontologyportal.org.

[4] http://wordnet.princeton.edu/.

set), which is regarded as the building block in WordNet. Each synset in WordNet is linked with other synsets through various kinds of paradigmatic semantic and lexical relations. It is noted that paradigmatic semantic relations (e.g., hypernymy, antonymy) are defined between *concepts* (i.e., synsets), while lexical relations (e.g., gradation) are defined between *words*. Table 1 shows the basic relations statistics of WordNet.

Table 1. WordNet 1.6 lexical semantic relations statistics

WN code	Relation's name	Count	POS
!	Antonymy	24608	N, V, Adj, Adv
@	Hypernym	78446	N, V
~	Hyponym	78446	N, V
#m	Member holonym	11849	N
%m	Member meronym	11849	N
#s	Substance holonym	709	N
%s	Substance meronym	709	N
#p	Part holonym	6883	N
%p	Part meronym	6883	N
*	Entailment	427	V
		122922	

As shown, WordNet contains mostly paradigmatic relations, i.e., relations among synsets with words belonging to the same part-of-speech. It is noted that from WordNet 2.0, the cross-POS links called **morpho-semantic links** have been introduced. This extension specifies the relations among words (synset members) that are semantically similar and morphologically related (Miller and Fellbaum 2003). Most of these links connect words with different syntactic classes (noun-verb, verb-adjective, noun-adjective). Currently, there are about tens of thousands of manually encoded connections, linking derivationally related words. The example is shown in (2).

(2) Manually encoded connections among derivationally related words
abandon\#v1 - abandonment\#n3
rule\#v6 - ruler\#n1
catch\#v4 - catcher\#n1

However, what WordNet does not currently inform about is the semantic nature of these relations. For example:

(3) Semantic relation
abandonment\#n3 is the EVENT of abandon\#
ruler\#n1 is the INSTRUMENT of rule\#v6
catcher\#n1 is the AGENT of catch\#v4

In order to make the semantic nature of the added morpho-semantic links more explicit, which is required for many NLP tasks, many studies have been proposed and implemented in wordnets of different languages like Czech (Pala et al. 2007), Zulu, Bulgarian and Serbian (Koeva et al. 2008). Table 2 shows the semantic classification of *-er* noun and verb pairs in Fellbaum et al. (2007), with the number of pairs given in the right-hand column.

Table 2. Semantic roles of *-er*

semantic roles of *-er*	number
Agent	2,584
Instrument	482
Inanimate agent/Cause	302
Event	224
Result	97
Undergoer	62
Body part	49
Purpose	57
Vehicle	36

3.2 Toward a multilingual LSRs knowledge base

3.2.1 Background

My initial motivation for this study can be traced back to the observation that, in the context of globalization, language resource goes global, too. The visioned new generation of language resource has taken its shape from static, closed and locally developed resources, to shared and distributed language services. Language resources reside over distributed places, and are choreographed by agents presiding the actions that can be executed over them, such as querying, collaborative development and validation, cross-resource integration and exchange of information. This is a long-term scenario based on content interoperability standards, sovran-national cooperation and development of accessible architectures enabling accessibility (Huang 2006).

Under this background, different strategies have been proposed to *multilingual LSRs acquisition.* However, copying or simple *porting* LSRs from one wordnet to another could possibly lead to invalid relations in the target wordnet. Tufiş and Cristea (2002) conjectured thus the Hierarchy Preservation Principle (HPP) for the automatic import of most of the semantic relations from Princeton WordNet into the Romanian wordnet.

According to Tufiş and Cristea (2002)'s explanation, HPP works as follows: if two synsets *S1_source* and *S2_source* in the source wordnet are connected by a semantic relation R, and assume that *S1_target* and *S2_target* are the correspondingly aligned synsets in the target wordnet, then they will be linked by the relation R. If there are intervening synsets between *S1_target* and *S2_target* in the source wordnet, then the relation R between the corresponding target synsets will be set up only if R is declared as *transitive* (i.e., R+, unlimited number of compositions, e.g. hypernym) or *partially transitive relation* (i.e., Rk with k a user-specified maximum number of compositions, larger than the number of intervening synsets between *S1_target* and *S2_target*.

However, this approach presumes the synonomous correspondence among pairs like < *S1_source, S1_target*>, < *S2_source, S2_target*>, and so on, which results in the neglect of the complex algebraic properties of LSRs.

3.2.2 (Naive) bootstrapping CWN from PWN/EWN

With these considerations in mind, in building Chinese Wordnet (CWN), our initial concentration was on the identification and definition of Chinese word senses, mapping them with Princeton WordNet synsets, and a very small subset of lexical semantic relations (LSRs) marked. What needs to be carried out fully is the comprehensive annotation of LSRs. It is presumed that essences of LSRs are *more* universal than word senses in human languages. So one way that I have been experimenting with is to bootstrap from Princeton WordNet (PWN) through the ECTED (English-Chinese Translation Equivalents Database) (Huang et al. 2002). A new set of inference rules has been devised to infer/extract LSRs automatically rather than manually. My preliminary works have shown that bootstrapping approach can not only enhance the shared upper lexical knowledge representation but also retain conceptual specificities in cross-cultural settings (Hsieh et al. 2006).

By bootstrapping PWN/EWN LSRs into CWN, I have launched the experiments based on the methods mentioned. The scope covers the over 10,000 senses currently proposed in CWN, and yields 36271 predicted relations. An initial evaluation by taking 40% of the whole relations yields 88% accuracy. As one would expect, some of the lexical relations (such as derivative, participle, region domain) are mostly valid cross-linguistically, which boils down to the important evaluation issue to be dealt with: Which lexical relations turn out to be not subject to automatic importing and would require human validations?

If a solid LSR algebraic foundation could be built through PWN/EWN, which can efficiently facilitate the checking of irregularities, then the bootstrapping approach based on these should be a reasonable extension in constructing prototypical multilingual wordnets. Figures 1 and 2 schematize the process. For example, given the rules 1 → 2 and 2 → 3, we can automatically get 1 → 4 on the local side, and 1 → 3 on the global side as well. For the languages with scare resources (say, XWN), it also saves time and effort in constructing wordnet-like resources.

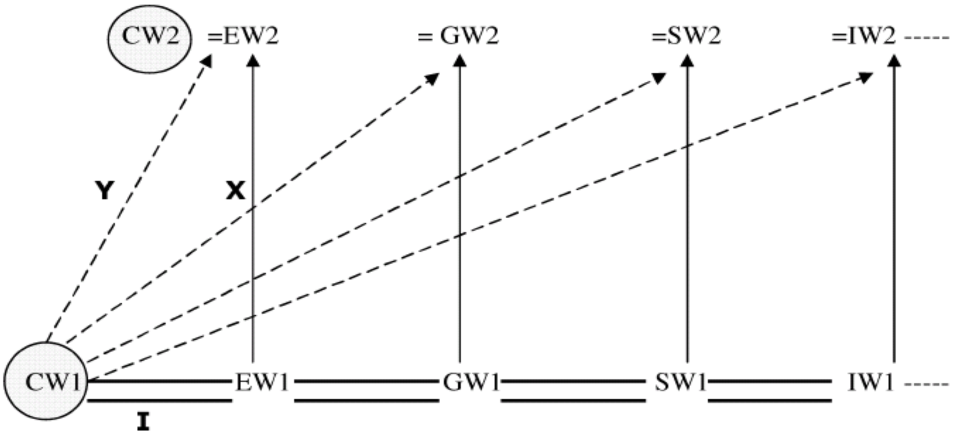


Figure 1. Multilingual wordnets bootstrapping (1)

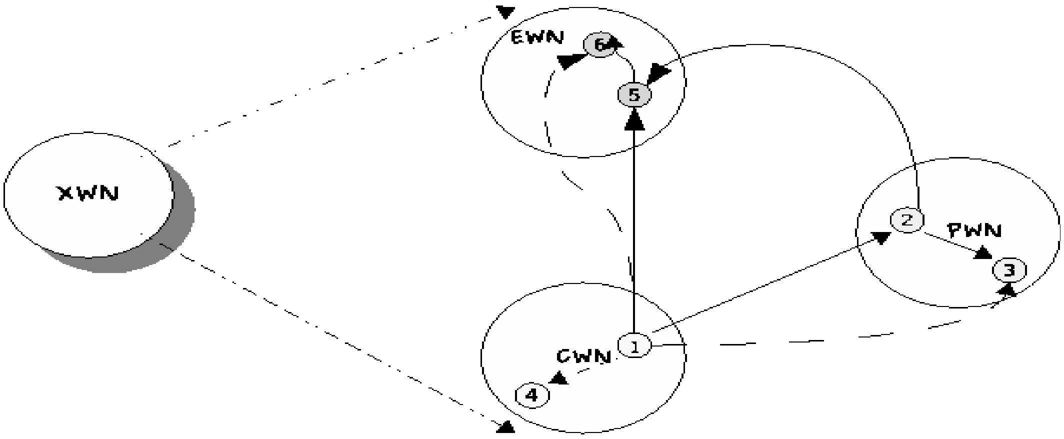


Figure 2. Multilingual wordnets bootstrapping (2)

This paper continues with the line of that research and provides a formal analysis supported by empirical data from PWN/EWN/CWN.

4. A formal approach to LSRs

Historically, the development of formal logic has been closely connected with the development of formal linguistics. Linguists usually employ set-theoretic models for formal semantics of natural languages. However, as Priss and Old (2004) proposed, semantics of knowledge representation formalisms can also be first interpreted as an algebra, which then further can be interpreted as a set-theoretic structure.

To begin with, some definitions of basic algebraic notions are provided.

(4) Definition 1 (*Relation*)

Let A and B be sets, a ***relation R*** *from A to B is a subset of* **A** × **B**, *the Cartesian product of* **A** *and* **B**. *If (a,b) ∈ R, we write aRb and say that "a is in relation R to b". A relation R on set A may have some of the following properties:*

– ***R*** *is* ***reflexive*** *if aRa for all a ∈ A.*
– ***R*** *is* ***symmetric*** *if aRb implies bRa for all a,b ∈ A.*
– ***R*** *is* ***antisymmetric*** *if aRb and bRa imply a = b for a,b ∈ A.*
– ***R*** *is* ***transitive*** *if aRb and bRc imply aRc for all a,b,c ∈ A.*

4.1 Algebraic structure of wordnet's LSRs

It has been observed that there are certain algebraic properties among paradigmatic lexical relations. For example, a lexical relation is said to be *transitive* if the fact that it holds between two elements A and B, and also between B and some third element C, guarantees that it holds between A and C (Cruse 1986).

4.2 Lexical relational algebraic structure

Our previous study (Huang et al. 2002) proposed a broader view on the *underlying inference logic* of LSRs by stipulating 26 rules. For example, given *i* = *HYP*, the following formula is provided.

(5) *i*= *HYP* formula

a. IF x = ANT
LSR y = HYP + ANT = ANT # CW2 is the antonym of CW1.
b. IF x = HYP
LSR y = HYP + HYP = HYP # CW2 is the hypernym of CW1.
c. IF x = NSYN
LSR y = HYP + NSYN = HYP # CW2 is the hypernym of CW1.
d. IF x = HOL
LSR y = HYP + HOL = HOL # CW2 is the holonym of CW1.
e. IF x = all other LSR's
LSR y = HYP + all other LSR's = ? #(Undecided)

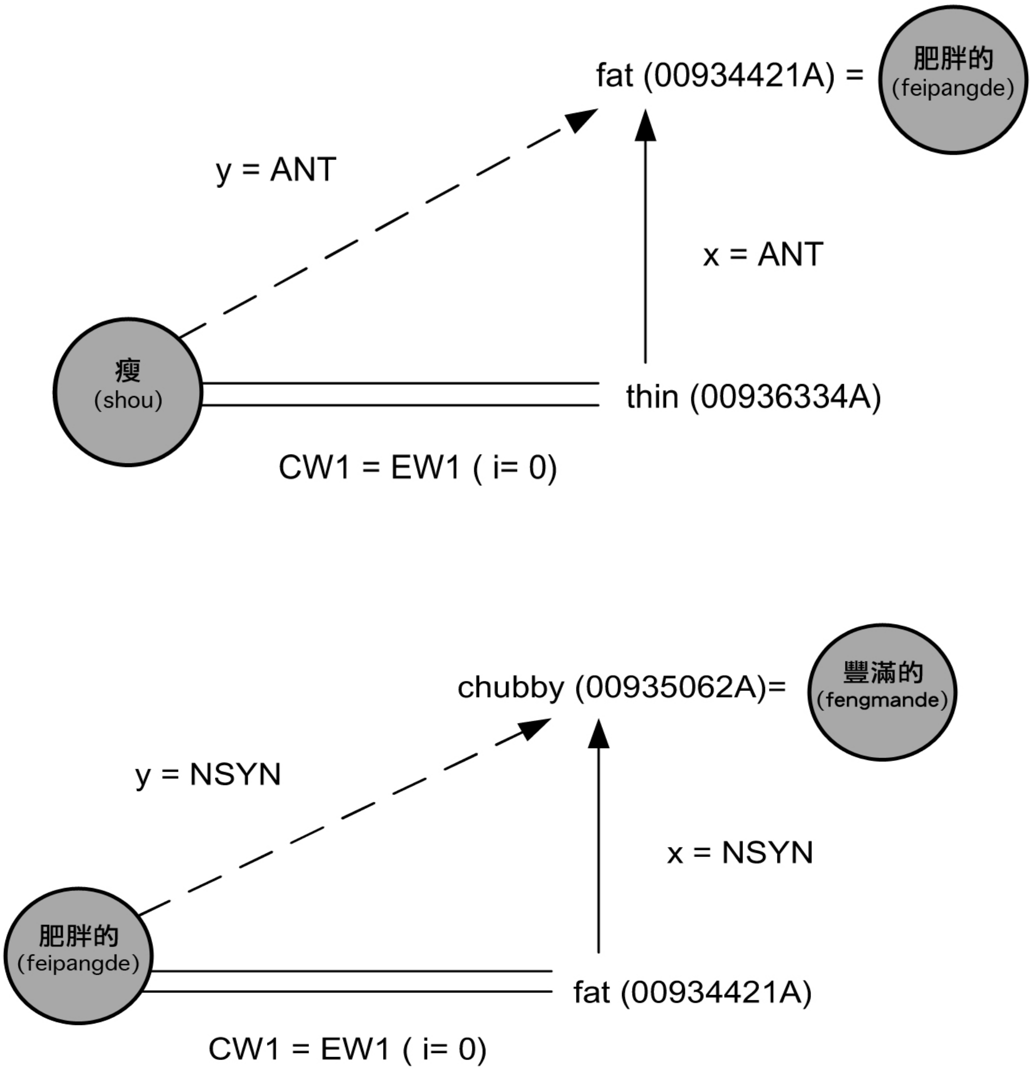


Figure 3. Translation-mediated LSR (when TEs are synonymous): The unknown LSR $y=0+x=x$

Similarly, given *i=HPO*, we can have the following rules:

(6) Rules of *i=HPO*

a. IF x = ANT
LSR y = HPO + ANT = ANT # CW2 is the antonym of CW1.
b. IF x = HPO
LSR y = HPO + HPO = HPO # CW2 is the hyponym of CW1.
c. IF x = NSYN
LSR y = HPO + NSYN = HPO # CW2 is the hyponym of CW1.
d. IF x = MER
LSR y = HPO + MER = MER # CW2 is the meronym of CW1.
e. IF x = all other LSR's
LSR y = HPO + all other LSR's = ? # (Undecided)

Figure 3 schematizes this model given i = synonymy. The complete set of inference rules is listed in Table 3.

Table 3. A set of inference rules of LSR's

	I	X	Y	Bootstrapped Results
1	HYP	ANT	ANT	{CW1, ANTONYM, CW2}
2	HYP	HYP	HYP	{CW1, HYPONOMY, CW2}
3	HYP	NSYN	HYP	{CW1, HYPONYM, CW2}
4	HYP	HOL	HOL	{CW1, HOLONYM, CW2}
5	HYP	all other LSRs undecided?		
6	HPO	ANT	ANT	{CW1, ANTONYM, CW2}
7	HPO	HPO	HPO	{CW1, HYPONYM, CW2}
8	HPO	NSYN	HPO	{CW1, HYPONYM, CW2}
9	HPO	MER	MER	{CW1, MERONYM, CW2}
10	HPO	all other LSRs undecided?		
11	NSYN	ANT	ANT	{CW1, ANTONYM, CW2}
12	NSYN	HYP	HYP	{CW1, HYPERNYM, CW2}
13	NSYN	HPO	HPO	{CW1, HYPONYM, CW2}
14	NSYN	NSYN	NSYN	{CW1, NEAR-SYNONYM, CW2}
15	NSYN	MER	MER	{CW1, MERONYM, CW2}
16	NSYN	HOL	HOL	{CW1, HOLONYM, CW2}
17	HOL	ANT	ANT	{CW1, ANTONYM, CW2}
18	HOL	HYP	HYP	{CW1, HYPONYM, CW2}

19	HOL	NSYN	HOL	{CW1, HOLONYM, CW2}
20	HOL	HOL	HOL	{CW1, HOLONYM, CW2}
21	HOL	all other LSRs undecided?		
22	MER	ANT	ANT	{CW1, ANTOMY, CW2}
23	MER	HPO	HPO	{CW1, HYPONYM, CW2}
24	MER	NSYN	MER	{CW1, MERONYM, CW2}
25	MER	MER	MER	{CW1, MERONYM, CW2}
26	MER	all other LSRs undecided?		

To ascertain that this bootstrapping strategy is theoretically sound and computational implementable, the logical speculation of the inference rules need strict proof and/or more statistical evidences. For example, I could hardly find instances in PWN where << *CW*1@*EW*1 >, < EW1#s EW2 >> => < EW2#s CW1> through such inference seems instinctive.[5]

This motivates me to propose further the notion of **LSRs Algebra**. The LSRs Algebra proposed here covers not only the traditional algebraic property already known as Transitivity (e.g., HYPO; MERO) and Asymmetricality (e.g. HYPO; MERO), but also covers the lexical semantic distribution of whole relation network by extending the previous inference rules. By this, new rules with significance might be extracted, such as << EW1(change state) TROPONYMY EW2 (thin) >, < EW2 (thin) ANTONOMY, EW3(thicken) >> => < EW1 TROPONYMY EW3 >. I will argue that a *lattice* theoretical approach to LSRs is a suitable starting point for formalizing them.

4.3 LSRs modeling proposed via Formal Concept Analysis

In the following, I will introduce the main ideas of *Formal Concept Analysis* method, a formal way that has been used to model *conceptual hierarchies* in terms of a special case of *lattice*, which is a special kind of *partial order relation*.

(7) Definition 2 (*Partial Order Relation*)

A reflexive, antisymmetric, and transitive relation R on a set A is called a ***partial order (relation)****. In this case, (A, R) is called a partially ordered set or poset.*

---

[5] The relation code used here can be referred to Table 1.

(8) Definition 3 (*Lattice*)

A ***lattice*** *is a structure consisting of a set A, a partial order relation* $\leq$*, and two binary operators* $\cap$ *(meet; intersection) and* $\cup$ *(join; union), which satisfy the following laws for all* $x, y, z \in L$*:*

–(L1: communicative): $x \cap y = y \cap x$, $x \cup y = y \cup x$;

–(L2: associative): $x \cap (y \cap z) = (x \cap y) \cap z$, $x \cup (y \cup z) = (x \cup y) \cup z$;

–(L3: absorption): $x \cap (x \cup y) = x$, $x \cup (x \cap y) = x$.

Two applications of (L3), namely, $x \cap x = x \cap (x \cup (x \cap x)) = x$*, lead to the additional law:*

–(L4: idempotent): $x \cap x = x$, $x \cup x = x$.

The Formal Concept Analysis (hereafter FCA) is a theory of data analysis in the field of applied mathematics, which is based on the mathematization of *concept* and *conceptual hierarchy*. It was introduced by a German mathematician Rudolf Wille in 1982.[6] Since it can identify *conceptual structures* among data sets, it has been successfully applied to a broad variety of domains such as sociology, medicine, computer science and industrial engineering.

The FCA method focuses on the *Concept Lattice Structures*, also called *Galois lattices*, arisen from binary data tables, which have been shown to provide a theoretical framework for a number of practical problems in information retrieval, software engineering, as well as knowledge representation and management. One of its best features is its capability of producing graphical visualizations of the inherent structures among data. Due to this capability, it can also be used as a fit tool in formalizing, revising and refining lexical databases, thesauri and ontologies.[7]

Priss (2005) proposed that FCA as a methodology of data analysis and knowledge representation has the potential to be applied to various linguistic problems. For instance, we can use FCA to build a lexical database, thesaurus or ontology, visualize conceptual structures in a lexical database, analyze semantic relations and identify inconsistencies among semantic relations in a lexical database.

To allow a mathematical description of extensions and intensions, FCA starts with the definition of a *formal context*.

---

[6] The introductory part is mainly based on Wolff (1993). For a more mathematical treatment of some of the topics covered here, the reader is referred to Ganter and Wille (1999). A lot of relevant publications can be found under http://www.mathematik.th-darmstadt.de/ags/, in both English and German.

[7] See Priss (1998) for an analysis for WordNet and Old (2002) for Roget's Thesaurus.

(9) Definition 4 (*Formal Context*)

A formal context is a triple K := (G,M,I), consisting of two sets G and M, and a binary relation I between G and M. That is $\boldsymbol{I \subseteq G \times M}$*. The elements of G and M are called* ***objects (Gegenstände)*** *and* ***attributes (Merkmale)****, respectively. The relation is written as gIm or* $(g,m) \in I$ *and is read as "the formal object g has the attribute m".*

A formal context can be represented by a *cross table* that has a row for each object *g*, a column for each attribute *m*, a cross in the row of *g* and the column of *m* of *gIm*.

(10) Definition 5

For $A \subseteq G$*, we define*

$A' := \{m \in M \mid \forall g \in A: (g,m) \in I\}$

and, analogously, for $B \subseteq M$*,*

$B' := \{g \in G \mid \forall m \in B: (g,m) \in I\}$

So in Table 4, A' {bus} = {four-tires plus, public, oil-burning} and B' {four-tires plus} = {car, train, bus} both hold.

(11) Definition 6 (*Formal Concept*)

A pair (A, B) is a formal concept C of the formal context (G, M, I) if and only if $A \subseteq G$*,* $B \subseteq M$*,* $A' = B$*, and* $A = B'$*.*

For a formal concept *C* := (*A*, *B*), A is called the *extent* (denoted by *Ext(c)*) and B is called the *intent* (denoted by *Int(c)*) of the formal concept. In the example of Table 4, ({car, bicycle, motorbike}, {private}) is a formal concept because A' {car, bicycle, motorbike} = {private}, and B' {private} = {car, bicycle, motorbike}.

The set of all formal concepts of a context K with the order relation ≤, denoted by *B* (*K*) (or *B* (G, M, I)), is called the **concept lattice** of *K*. It is always a complete lattice, i.e. for each subset of concepts, there is always a unique greatest common subconcept and a unique least common superconcept. Figure 4 shows the concept lattice of the formal context in Table 4 in the form of a line diagram.

Concept lattices can be depicted as *line diagrams* as in Figure 4, in which a formal concept is represented by a small circle. For each formal object *g*, the smallest formal concept to whose extent *g* belongs is denoted by γg; and for each formal attribute *m*, the largest formal concept to whose intent *m* belongs is denoted by μm. The concepts γg and μm are called *object concept* and *attribute concept*, respectively. In the line diagram it is not necessary to include either the full extent or intent for each concept;

instead, the name (verbal form) of each formal object *g* is written slightly above the circle of μm.

In a line diagram, the extent of a formal concept consists of all objects whose labels are attached to subconcepts. Analogously, the intent consists of all attributes attached to superconcepts. For example, the concept labeled *oil-burning* has {car, ambulance, motorbike, bus} as extent, and {oil-burning, two-tires} as intent. Based on that, FCA method can be useful in concept learning if we add more objects and attributes. Figure 5 shows a more complex concept lattice of the formal context by adding more objects.

The most important structure on *B* (G, M, I) is given by the *subconcept-superconcept relation* that is defined by

$$(A_1, B_1) \leq (A_2, B_2) :\Leftrightarrow A_1 \subseteq A_2 (\Leftrightarrow B_2 \subseteq B_1).$$

For example, in Table 4, ({car, bicycle, motorbike}, {private}) as a formal superconcept of ({motorbike}, {four-tires minus, private, oil-burning}), has more objects but fewer attributes than ({motorbike}, {four-tires minus, private, oil-burning}).

Table 4. A formal context of vehicles

	two-tires	four-tires plus	public	private	oil-burning
vehicle					
car		✓		✓	✓
train		✓	✓		
bicycle	✓			✓	
ambulance		✓	✓		✓
motorbike	✓			✓	✓
bus		✓	✓		✓

It follows from this definition that each formal concept is a formal subconcept of itself, in contrast to the natural language use of *subconcept*, which precludes a concept from being a subconcept of itself. The relation ≤ is a mathematical order relation called *formal conceptual ordering* on *B* (G, M, I) with which the set of all formal concepts forms a mathematical lattice denoted by *B* (G, M, I).

Until now, I have only illustrated the fomalization of *attribution relation* (i.e., the *has-a* relation between objects and attributes in a formal context). In the WordNet-like settings, other LSRs can also be formalized in similar ways.

4.4 Fuzzy LSRs and Concept Lattice

The Concept Lattice has some other advanced features over other representations.[8] However, the basic setting of FCA is well-suited for attributes which are crisp, i.e., each object either has or does not have the attribute. In my bootstrapping experiments, human evaluation shows that many predicted LSRs are fuzzy rather than crisp. That is, in many cases, it is a matter of degree to which a synset is *lexical-semantically related* to another synset.

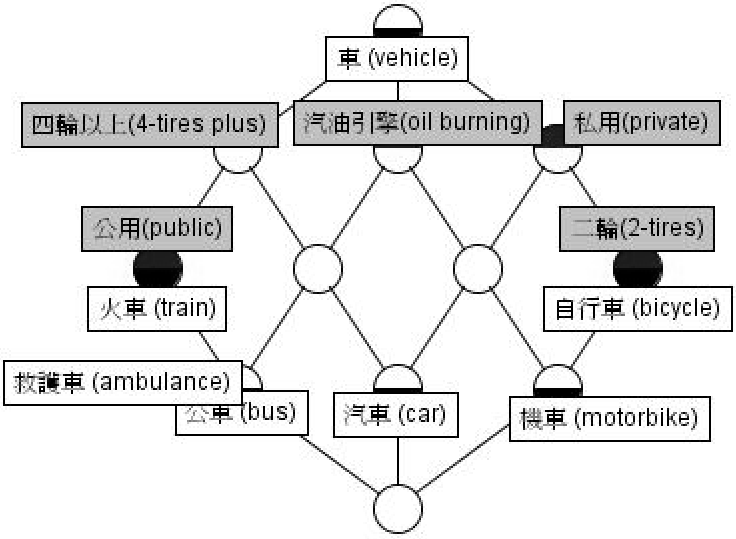


Figure 4. A concept lattice represented by a line diagram

---

[8] More details of the FCA can be found in Ganter and Wille (1996).

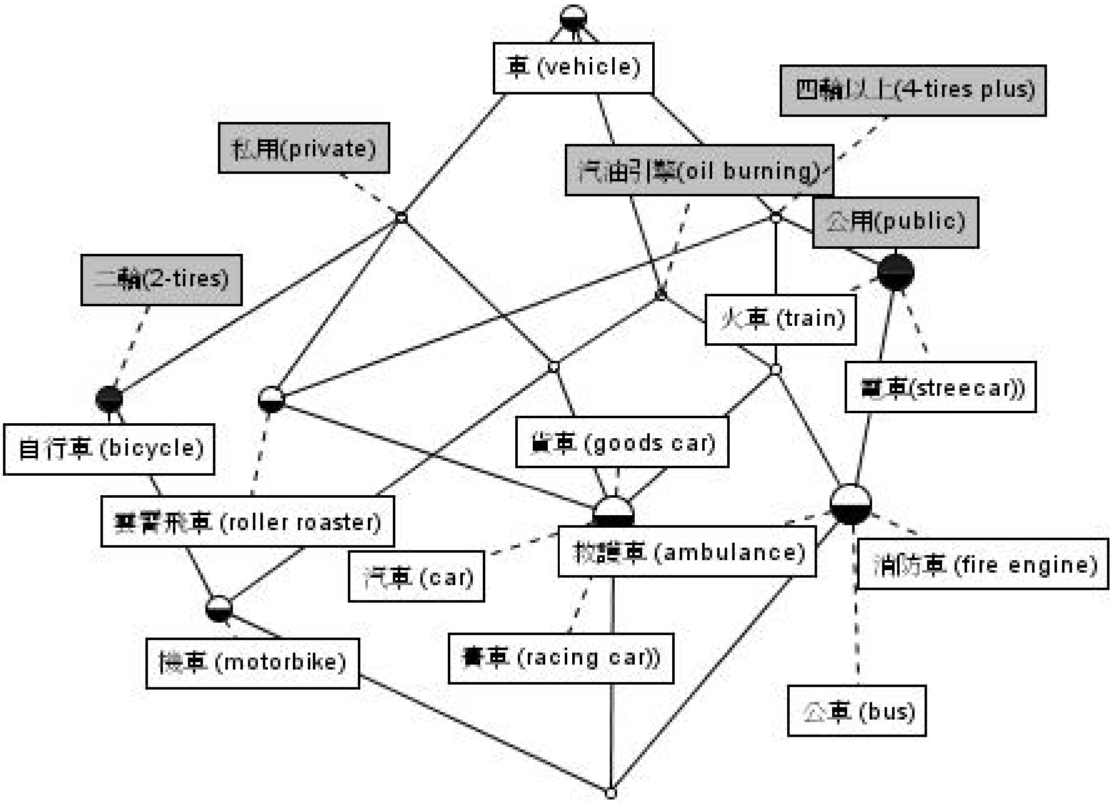


Figure 5. A more complex concept lattice

Therefore, in addition to logical considerations, I have also introduced the notion of *fuzzy lexical relations* as the solid empirical foundation of the proposed LSR algebra. That is, the bootstrapped LSRs are coupled with weights expressing the strength of associations based on the ration calculated from Princeton WordNet (i.e., degrees are taken from the scale *L* (WordNet) of truth degrees. By putting FCA in fuzzy setting, LSRs get no longer a binary value, but a membership value between 0 and 1. Table 5 shows a bootstrapped table with probabilities marked.

This extended FCA model can ease the task of automatic evaluation of LSRs. When several relations derived in the integration process have confidence scores greater than certain thresholds, the predicted LSRs are to be determined, and can be piped into manual verification. The results can also be compared with human evaluation. For the ease of browsing and manual checking, we also developed a visualization interface (Figure 6; on the next page).[9]

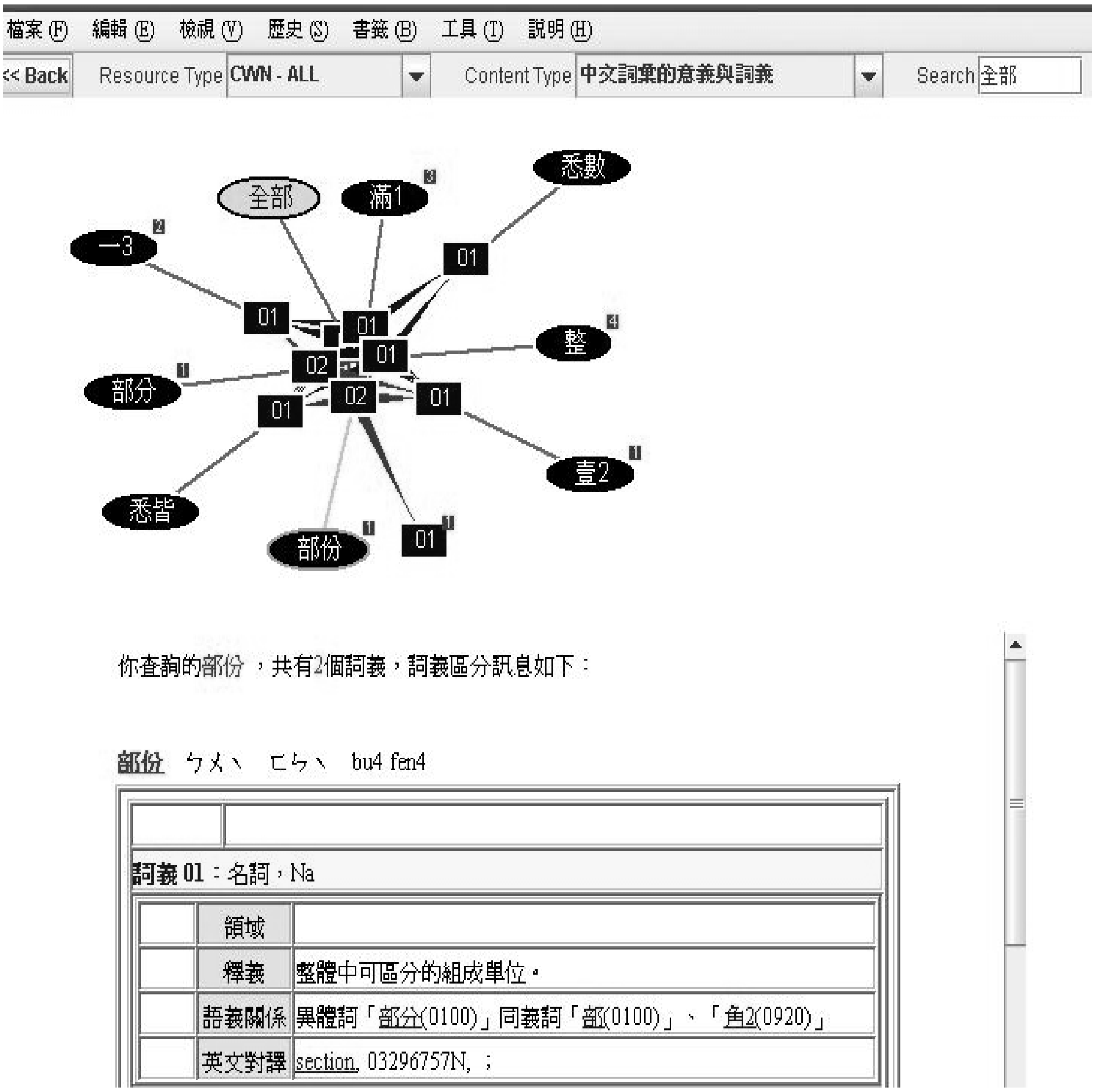
---

[9] http://cwn.ling.sinica.edu.tw/cwnviz/, still under construction.

Table 5. PWN-CWN LSRs bootstrapping table

PWN-HYPONYMY	synset1	synset2	synset3	synset4	synset5
synset1		↗			
synset2				↙	
synset3					
synset4	↗		↗		
synset5					

CWN-HYPONYMY	synset1	synset2	synset3	synset4	synset5
synset1		↗			
synset2				↙(0.80)	
synset3					
synset4	↗(0.45)		↗		
synset5					



Resource Type CWN - ALL

Content Type 中文詞彙的意義與詞義

Search 全部

你查詢的部份 ，共有2個詞義，詞義區分訊息如下：

部份 ㄅㄨˋ ㄈㄣˋ bu4 fen4

詞義 01：名詞，Na

領域	
釋義	整體中可區分的組成單位。
語義關係	異體詞「部分(0100)」同義詞「部(0100)」、「角2(0920)」
英文對譯	section, 03296757N, ；

Figure 6. Visualization of LSRs

5. Conclusion

Lexical semantic relations offer rich linguistic and conceptual knowledge information and are the most to fill in for wordnets. They are likely to be similarities that can be duplicated from language to language. Since LSRs represent complex knowledge, I assume that knowledge from different languages encoded in wordnets also tends to compliment each other. On closer inspection based on my previous experiments, however, one discovers that such formalization is quite challenging to obtain; it may be difficult to design a formalization that is both mathematically appropriate and has a semantics which matches a linguist's intuition.

In this paper, I propose a mathematical formalization of wordnet-driven LSRs through the Formal Concept Analysis in the fuzzy settings. Though the FCA method does not provide a complete axiomatic system for LSRs, it can facilitate the investigation of the logic properties of these relations, and discover the irregularities in the implementation of LSRs in multi-lingual lexical database. In addition, by extending Concept Lattice to Fuzzy Concept Lattice, the automatic bootstrapped repository of multilingual LSRs will become a crucial workable language resource for relevant research fields. The process of formalizing LSRs with formal concept analysis methods should be able to highlight problems and serve as a foundation for solutions that can support linguistic theoretical explorations and applications for multilingual NLP tasks in the future.

References

Atserias, Jordi, Luís Villarejo, German Rigau, Eneko Agirre, John Carroll, Bernardo Magnini, and Piek Vossen. 2004. The MEANING multilingual central repository. *Proceedings of the Second International Global WordNet Conference*, 23-30. Brno: Masaryk University.

Baldwin, Timothy. 2006. Learning lexical semantic representations. Lecture notes from the ACL/HCSNet Advanced Program in Natural Language Processing. University of Melbourne, Melbourne, Australia.

Bosch, Sonja, Christiane Fellbaum, and Karel Pala. 2008. Enhancing WordNets with morphological relations: A case study from Czech, English and Zulu. *Proceedings of the Fourth Global WordNet Conference*, 74-90. Szeged: University of Szeged, Department of Informatics.

Climent, Salvador, Horatio Rodriguez, and Julio Gonzalo. 2006. Definition of the links and subsets for nouns of the EuroWordNet project. *EuroWordNet Project*.

Amsterdam: University of Amsterdam.

Cruse, D. Alan. 1986. *Lexical Semantics*. London: Cambridge University Press.

Dahlberg, Ingetraud. 1995. Conceptual structures and systematization. *International Forum on Information and Documentation* 20.3:9-24.

DIN 2330. 1979. *Begriffe und Benennungen: Allgemeine Grundsätze [Concepts and Terms: General Principles]*. Berlin-Köln: Beuth Verlag.

DIN 32705.1987. *Klassifikationssyteme [Classification System]*. Berlin-Köln: Beuth Verlag.

Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.

Fellbaum, Christiane, Anne Osherson, and Peter E. Clark. 2007. Putting semantics to WordNet's "morphosemantic" links. *Proceedings of the Third Language and Technology Conference*, 226-230. Poznań: Wydawnictwo Poznańskie.

Ganter, Bernhard, and Rudolf Wille. 1996. Applied lattice theory: Formal concept analysis. *General Lattice Theory*, ed. by George A. Grätzer and B. A. Davey, 591-606. Basel: Birkhäuser Verlag.

Ganter, Bernhard, and Rudolf Wille. 1999. *Formal Concept Analysis: Mathematical Foundations*. Berlin: Springer.

Hsieh, Shu-kai, Sîmon Petr, and Chu-ren Huang. 2006. Large-scale acquisition of lexical semantic relations: A preliminary study on Chinese Wordnet. Paper presented at ROCLING XVIII: Conference on Computational Linguistics and Speech Processing, Academia Sinica.

Huang, Chu-ren, I-ju E. Tseng, and Dylan B. S. Tsai. 2002. Translating lexical semantic relations: The first step toward multilingual wordnets. *Proceedings of COLING-02: SEMANet: Building and Using Semantic Networks*, 2-8. Taipei: Academia Sinica.

Huang, Chu-Ren, Elanna I. J. Tseng, Dylan B. S. Tsai, Brian Murphy. 2003. Cross-lingual portability of semantic relations: Bootstrapping Chinese Wordnet with English WordNet relations. *Language and Linguistics* 4.3:509-532.

Huang, Chu-Ren, I-Li Su, Jia-fei Hong, and Xiang-bin Li. 2005. Cross-lingual conversion of lexical semantic relations: Building parallel wordnets. *Proceedings of the Fifth Workshop on Asian Language Resources and First Symposium on Asian Language Resources Network*, 48-55. N.p.: Asian Federation of Natural Language Processing.

Huang, Chu-ren. 2006. Towards global Wordnet grid: Infrastructure for language technology in the age of multilingualism. Keynote speech at ROCLING XVIII: Conference on Computational Linguistics and Speech Processing, Academia Sinica.

Koeva, Svetla, Cvetana Krstev, and Duško Vitas. 2008. Morpho-semantic relations in WordNet: A case study for two Slavic languages. *Proceedings of the Fourth Global WordNet Conference*, 239-254. Szeged: University of Szeged, Department of Informatics.

Miller, A. George, and Christiane Fellbaum. 2003. Morphosemantic links in WordNet. *Journal of Traitement Automatique de Langue* [*Journal of Natural Language Processing*] 44.2:69-80.

Old, L. John. 2002. Information cartography applied to the semantics of Roget's Thesaurus. *Proceedings of the 13th Midwest Artificial Intelligence and Cognitive Science Conference*. Chicago: Illinois Institute of Technology. http://www.dcs.napier.ac.uk/~cs171/LJOld/papers/Semantic%20Cartography.pdf.

Pala, Karel, and Dana Hlaváčková. 2007. Derivational relations in Czech WordNet. *Proceedings of the Workshop on Balto-Slavonic*, 75-81. Prague: The Association for Computational Linguistics.

Pianta, Emanuele, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: Developing an aligned multilingual database. *Proceedings of the First International Conference on Global WordNet*, 293-302. Mysore: Central Institute of Indian Languages.

Priss, Uta. 1996. *Relational Concept Analysis: Semantic Structures in Dictionaries and Lexical Databases*. Doctoral dissertation, Technische Universität Darmstadt.

Priss, Uta. 1998. The formalization of WordNet by methods of relational concept analysis. *WordNet: An Electronic Lexical Database and Some of its Applications*, ed. by Christiane Fellbaum, 179-196. Cambridge, Massachusetts: The MIT press.

Priss, Uta. 2003. Linguistic applications of formal concept analysis. *Proceedings of ICFCA*, 234-243. Darmstadt: Darmstadt University of Technology.

Priss, Uta, and John L. Old. 2004. Modeling lexical databases with formal concept analysis. *Journal of Universal Computer Science* 10.8:967-984.

Priss, Uta. 2005. Linguistic applications of formal concept analysis. *Formal Concept Analysis, Foundations and Applications*, ed. by Bernhard Ganter, Gerd Stumme, and Rudolf Wille, 149-160. Heidelberg: Springer Verlag.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, Massachusetts: The MIT press.

Saussure, Ferdinand de. [1916] 1983. *Cours de Linguistique Générale* [*Course in General Linguistics*]. Trans. by Roy Harris. London: Duckworth.

Tufiş, Dan, and Dan Cristea. 2002. Methodological issues in building the Romanian Wordnet and consistency checks in Balkanet. *Proceedings of LREC2002 Workshop on Wordnet Structures and Standardisation*, 35-41. Las Palmas: University of Las Palmas de Gran Canaria.

Vossen, Piek. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer Academic Publishers.

Wolff, Karl Erich. 1993. A first course in formal concept analysis. *SoftStat'93: Advances in statistical software*, ed. by Frank Faulbaum, 429-438. Stuttgart: Gustav Fischer Verlag.

Xu, Ming-Wei, Jia-Fei Hong, Shu-Kai Hsieh, and Chu-Ren Huang. 2008. CWN-Viz: Semantic relation visualization in Chinese wordnet. *Proceedings of the Fourth Global WordNet Conference*, 506-519. Szeged: University of Szeged, Department of Informatics.

[Received 22 April 2008; revised 22 Dec 2008; accepted 6 January 2009]

Department of English
National Taiwan Normal University
Taipei, TAIWAN
Shu-Kai Hsieh: shukai@ntnu.edu.tw

詞彙語意關係的形式描述

謝舒凱

國立臺灣師範大學

詞彙語意關係在近年來的自然語言處理研究中扮演重要的角色，也同樣地影響著詞彙語意資源的建構。在此脈絡下，本文回顧了詞彙語意關係的研究，並利用擴展的形式概念分析提出一套詞彙語意關係的形式模組。作者認爲這個模組能突顯語意與文化差異的問題，同時也能支持詞彙理論上的解釋以及多語化的詞彙網路應用。

關鍵詞：詞彙語意學、計算詞彙、形式概念分析