# CWN-LMF: Chinese WordNet in the Lexical Markup Framework

**Lung-Hao Lee[1], Shu-Kai Hsieh[2], Chu-Ren Huang[1,3]**

[1]Institute of Linguistics, Academia Sinica
[2]Department of English, National Taiwan Normal University
[3]Department of Chinese & Bilingual Studies, The Hong Kong Polytechnic University
[1]128 Academia Road, Section 2, Taipei 115, Taiwan
[2]162 He-ping East Road, Section 1, Taipei 106, Taiwan
[3]Hung Hom, Kowloon, Hong Kong
[1]{lunghao,churen}@gate.sinica.edu.tw
[2]shukai@ntnu.edu.tw
[3]churen.huang@inet.polyu.edu.hk

## Abstract

Lexical Markup Framework (LMF, ISO-24613) is the ISO standard which provides a common standardized framework for the construction of natural language processing lexicons. LMF facilitates data exchange among computational linguistic resources, and also promises a convenient uniformity for future application. This study describes the design and implementation of the WordNet-LMF used to represent lexical semantics in Chinese WordNet. The compiled CWN-LMF will be released to the community for linguistic researches.

## 1 Introduction

Princeton WordNet[1] is an English lexical database that groups nouns, verbs, adjectives and adverbs into sets of cognitive synonyms, which are named as synsets (Fellbaum, 1998; Miller, 1995). The Global WordNet Association (GWA)[2] built on the results of Princeton WordNet and Euro WordNet (Vossen, 2004) is a free and public association that provides a platform to share and connect all languages in the world. For Mandarin Chinese in Taiwan, Huang et al. (2004) constructed the Academia Sinica Bilingual Ontological Wordnet (Sinica BOW) which integrates WordNet, English-Chinese Translation Equivalents Database (ECTED) and SUMO for cross-language linguistic studies. As a follow-up, Chinese WordNet (CWN) has been built as a robust lexical knowledge system which also embodies a precise expression of sense relations (Huang et al., 2008). In recent years, WordNet-like resources have become one of the most reliable and essential resources for linguistic studies for all languages (Magnini and Cavaglia, 2000; Soria et al. 2009; Strapparava and Valitutti, 2004).

Lexical Markup Framework (LMF, ISO-24613) is the ISO standard which provides a common standardized framework for the construction of natural language processing lexicons (Francopoulo et al., 2009). One important purpose of LMF is to define a standard for lexicons which covers multilingual lexical information (Francopoulo et al., 2006b). In this study, we describe the design and implementation of the Wordnet-LMF (Soria et al. 2009) to represent lexical semantics in Chinese WordNet.

The rest of this paper is organized as follows: Section 2 introduces Chinese WordNet and Lexical Markup Framework. Section 3 describes how we represent Chinese WordNet in the Lexical Markup Framework (CWN-LMF). Section 4 presents an example on Chinese word sense distinction using CWN-LMF format. Quantitative analysis of compiled CWN-LMF is presented in Section 5. We also describe the application scenario using CWN-LMF for information interoperability of lexical semantics in Section 6. Section 7 discusses the experience and difficulties of encoding CWN into Wordnet-LMF. Finally, Section 8 concludes this study with future research.

[1] Wordnet, available online at http://wordnetweb.princeton.edu/perl/webwn
[2] Global WordNet Association (GWA), available online at http://www.globalwordnet.org/



*Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009*, pages 123–130,
Suntec, Singapore, 6-7 August 2009. 

## 2 Related Work

### 2.1 Chinese WordNet

Creating a semantic relation-based language resource is a time consuming and labor intensive task, especially for Chinese due to the unobvious definition and distinction among characters, morphemes and words. Chinese WordNet [3] (CWN) has been built by Academia Sinica and is successively extended its scope so far. Lemmas included in CWN mainly fall on the medium frequency words. Each lexical entry is analyzed according to the guidelines of Chinese word sense distinctions (CKIP, 2003; Huang et al. 2003) which contain information including Part-of-Speech, sense definition, example sentences, corresponding English synset(s) from Princeton WordNet, lexical semantic relations and so on. Unlike Princeton WordNet, CWN has not been constructed mainly on the synsets and semantic relations. Rather it focuses to provide precise expression for the Chinese sense division and the semantic relations needs to be based on the linguistic theories, especially lexical semantics (Huang et al., 2008). Moreover, Huang et al. (2005) designed and implemented the Sinica Sense Management System (SSMS) to store and manage word sense data generated in the analysis stage. SSMS is meaning-driven. Each sense of a lemma is identified specifically using a unique identifier and given a separate entry. There are 8,646 lemmas / 25,961 senses until December 2008 have been analyzed and stored in SSMS. Figure 1 shows the result of sense distinction for 足跡 zu-ji 'footprint' as an example in Chinese WordNet.

Huang et al. (2004) proposed Domain Lexico-Taxonomy (DLT) as a domain taxonomy populated with lexical entries. By using DLT with Chinese WordNet and Domain Taxonomy, there were 15,160 Chinese senses that linked and distributed in 463 domain nodes. In addition, Huang et al. (2005) further applied DLT approach to a Chinese thesaurus called as CiLin and showed with evaluation that DLT approach is robust since the size and number of domain lexica increased effectively.

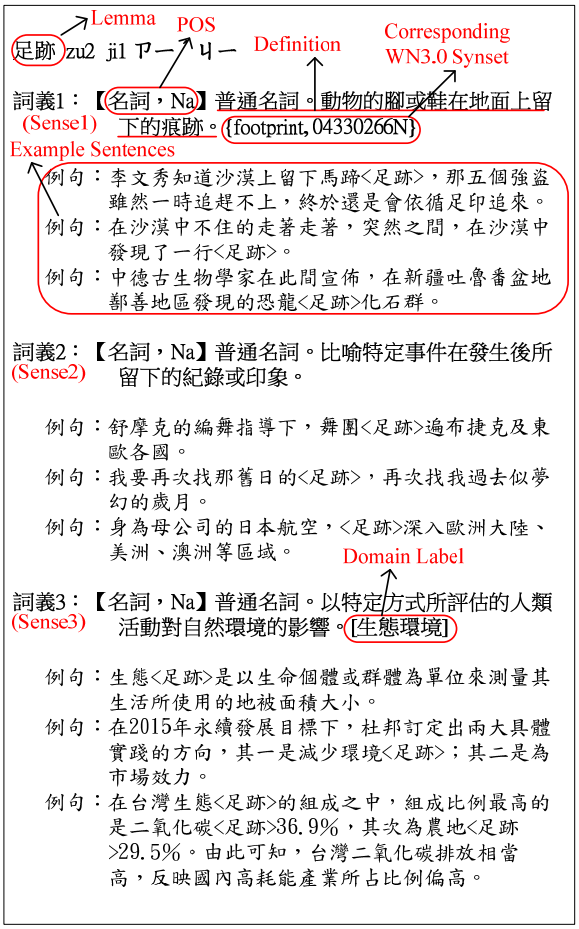


Figure1: The result of sense distinction for "zu2 ji1 (footprint)".

### 2.2 Lexical Markup Framework

Lexical Markup Framework (LMF, ISO-24613) is the ISO standard for natural language processing lexicons and machine readable dictionaries. The goals of LMF are to provide a common model for the creation and use of lexical resources, and to manage the exchange of data between them. Francopoulo et al. (2006a; 2009) offered a snapshot of how LMF represents multilingual lexicons. LMF facilitates data exchange among computational linguistic resources and also promises a convenient uniformity for future application. More updated information can be found online at http://www.lexicalmarkupframework.org .

Soria et al. (2009) proposed a Wordnet-LMF developed in the framework of the KYOTO [4] project as a standardized interoperability format for the interchange of lexico-semantic information. Wordnet-LMF is an LMF dialect tailored to encode lexical resources adhering to the Word-

[3] Chinese WordNet, available online at http://cwn.ling.sinica.edu.tw/

[4] KYOTO, available online at http://www.kyoto-project.eu/

Net model of lexical knowledge representation. Wordnet-LMF was designed by adhering to LMF principles yet taking into account on the one hand, the peculiarities of the Wordnet model, and on the other by trying to maximize the efficiency of the format.

If we take Princeton WordNet 3.0 synset {footprint_1} for example, a Wordnet-LMF representation can be found in Figure 2. The details will be explained in Section 3.

```
<Synset id="eng-30-06645039-n" baseConcept="1">
<Definition gloss="mark of a foot or shoe on a surface">
<Statement example="the police made casts of the
footprints in the soft earth outside the window"/>
</Definition>
<SynsetRelations>
<SynsetRelation target="eng-30-06798750-n"
relType="has_hyperonym">
</SynsetRelation>
<SynsetRelation target="eng-30-06645266-n"
relType="has_hyponym">
</SynsetRelation>
</SynsetRelations>
<MonolingualExternalRefs>
<MonolingualExternalRef externalSystem="Wordnet1.6"
externalReference="eng-16-01234567-n">
<MonolingualExternalRef externalSystem="SUMO"
externalReference="superficialPart" relType="at">
</MonolingualExternalRefs>
<Synset>
```

Figure 2: An example of Wordnet-LMF format.

## 3 CWN in the Lexical Markup Framework (CWN-LMF)

Wordnet-LMF is used to represent lexical semantics in Chinese WordNet. As *LexicalResource* is the root element in Wordnet-LMF, it has three children: one *GlobalInformation* element, one or more *Lexicon* elements, zero or *one SenseAxes* element. This means the object *LexicalResource* is the container for possibly more than one *lexicon;* inter-lingual correspondences are grouped in *SenseAxes* section. The details are presented as follows.

### 3.1 Global Information

The element named as *GlobalInformation* is used to describe general information about the lexical resource. The attribute "label" is a free text field. Example as follows:
<GlobalInformation label="Compile Chinese Wordnet entries using Wordnet-LMF">

### 3.2 Lexicon

In CWN-LMF, only one element *Lexicon* is used to contain a monolingual resource as a set of *LexicalEntry* instances followed by a set of *Synset* elements. The following attributes are specified:

- languageCoding: It has "ISO 639-3" as a fixed value.
- language: The standardized 3-letter language coding, e.g. zho, is used to specify the language represented by the lexical resource. It is a required attribute.
- owner: It is a required attribute to specify the copyright holder
- version: It is a required attribute to specify the resource version.
- label: It is used to record additional information that may be needed. This attribute is optional.

Example as follows:
<Lexicon languageCoding="ISO 639-3" label="Chinese WordNet 1.6" language="zho", owner="Academia Sinica", version="1.6">.

#### 3.2.1 Lexical Entry

A *LexicalEntry* element can contain one lemma and one sense and has an optional attribute "id" which means a unique identifier.

The element, Lemma, represents a word form chosen by convention to designate the lexical entry. It contains the following attributes:

- partOfSpeech: It is a required attribute. This attribute takes as its value the part-of –speech value that according to WordNet conventions is usually specified for a synset. There are four part-of-speech notations that are used in CWN-LMF. The notation "n" is represented as a noun; the notation "v" is represented as a verb; the notation "a" is represented as an adjective; the notation "r" is represented as an adverb; and the other POS tags are represented as "s".
- writtenForm: It is added in case that "id" of *LexicalEntry* is numerical and it takes Unicode strings as values. This attribute is optional.

The *Sense* element represents one meaning of a lexical entry. For WordNet representation, it represents the variant of a synset. Required attributes are:

- id: It must be specified according to the convention used in Chinese WordNet, i.e. word_sense#nr.. For example, "環境_1" means that the first sense of lemma 環境 huan-jing 'environment'.
- synset: It takes as its value the ID of the synset to which the particular sense of the variant belongs. The ID of the synset will be described in the next subsection.

Take the first sense of lemma 環境 huan-jing 'environment' for example, it will be represented as follows:

```
<LexicalEntry>
    <Lemma writtenForm="環境" partOfSpeech="n"></Lemma>
   <Sense id="環境_1" synset=" zho-16-06640901-n"></Sense>
</LexicalEntry>
```

### 3.2.2 Synset

This element encodes information about a Chinese WordNet synset. *Synset* elements can contain one *Definition*, optional *SynsetRelations* and *MonolingualExternalRefs* elements. Required attributes for *Synset* element are the following:

- id: It is a unique identifier. The agreed syntax is "languageCode-version-id-POS". For example, "zho-16-06640901-n" is unique identifier of the first sense of lemma 環境 huan-jing 'environment'.
- baseConcept: Values for the *baseConcept* attribute will be numerical (1,2,3), which correspond to the BaseConcept sets. If the sense belongs to the first-class basic words of NEDO project (Tokunaga et al. 2006), we encode it as 1. Similarly, if the sense belongs to second-class basic words, we encode it as 2. The other senses will be encoded as 3 if they are not basic words.

The element *Definition* allows the representation of the gloss associated with each synset in attribute "gloss". The required attribute "example" of the element *Statement* contains the examples of use associated with the synset .

*SynsetRelations* is a bracketing element for grouping all *SynsetRelation* elements. Relations between synsets are codified by means of *SynsetRelation* elements, one per relation. Required attributes are:

- target: It contains the ID value of the synset that is the target of the relation.
- relType: It means the particular type. There are nine semantic relations in Chinese WordNet, including "has_synonym", "has_nearsynonym", "has_hypernym", "has_hyponym", "has_holonym", "has_meronym", "has_paranym", "has_antonym" and "has_variant". Among them, the semantic relation paranymy is used to refer to relation between any two lexical items belonging to the same semantic classification (Huang et al. 2008). For example, the set of "spring/summer/fall/winter" has paranymy relation of main concept of "seasons in a year".

*MonolingualExternalRefs* is a bracketing element to group all *MonolingualExternalRef* elements. *MonolingualExternalRef* elements must be used to represent links between a Sense or Synset and other resources, such as an ontology, a database or other lexical resources. Attributes are:

- externalSystem: It is a required attribute to describe the name of the external resource. For instance, possible values are "domain" (Magnini and Cavaglia, 2000), "SUMO" (Niles and Pease, 2001), and "Wordnet 3.0" for recording SenseKey values.
- externalReference: It means the particular identifier or node. This attribute is required.
- relType: It is optional attribute. If the "externalSystem" is "SUMO". "relType" is the type of relations with SUMO ontology nodes. Possible values are "at", "plus", and "equal".

We use the first sense of lemma 環境 huan-jing ‘environment’ to illustrate as follows:

```
<Synset id="zho-16-06640901-n" baseCon-
cept="2">
  <Definition gloss="普通名詞。人類及其他
  生物生存的空間。">
    <Statement example="人類與非人類都非
    常脆弱，常因自然環境改變而受到嚴重
    傷害。"/>
  </Definition>
  <SynsetRelations>
    <SynsetRelation target="zho-16-
    07029502-n" relType="has_synonym">
    </SynsetRelation>
  </SynsetRelations>
  <MonolingualExternalRefs>
    <MonolingualExternalRef externalSys
    tem="SUMO" externalRefe
    rence="GeographicArea" rel
    Type="plus"/>
  </MonolingualExternalRefs>
</Synset>
```

### 3.3 SenseAxes

*SenseAxes* is a bracketing element that groups together *SenseAxis* elements used for inter-lingual correspondences. The *SenseAxis* element is a means to group synsets belonging to different monolingual wordnets and sharing the same equivalent relation to Princeton WordNet 3.0. Required attributes are:

- id: It is a unique identifier.
- relType: It specifies the particular type of correspondence among synsets belonging to different resources. We use “eq_synonym” to represent equal synonym relation between Chinese Wordnet and Princeton WordNet.

For instance, Chinese synset zho-16-06640901-n maps onto English synset eng-30-08567235-n by means of an eq_synonym relation. This will be represented as follows:

```
<SenseAxes>
  <SenseAxis id="sa_zho16-eng30_5709" rel
  Type="eq_synonym">
    <Target ID="zho-16-06640901-n"/>
    <Target ID="eng-30-08567235-n"/>
  </SenseAxis>
</SenseAxes>
```

## 4 An Example of CWN-LMF Format

Take 自然 zi-ran ‘nature’ as an example shown in Figure 3. 自然 has six senses (some of them are abridged in the figure). Id attribute of the first sense is 自然_1 and its synset is called “zho-16-03059301-n”. This encoding of synset stands for 自然_1 with the unique ID 03059301 in Chinese WordNet version 1.6 and its part-of-speech is noun. Moreover, one can also learn that 自然_1 has a synonym, 大自然_1 (zho-16-06653601-n). Meanwhile, this sense is also corresponded to IEEE SUMO. Finally, this compiled CWN-LMF version is pointed to Princeton WordNet 3.0, i.e. Chinese synset “zho-16-03059301-n” maps onto English synset “eng-30-11408559-n” by means of an eq_synonym relation.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE LexicalResource SYSTEM "kyoto_wn.dtd">
<LexicalResource>
  <GlobalInformation label="CWN-LMF" />
  <Lexicon languageCoding="ISO 693-3" label="Chinese
   Wordnet 1.6" language="zho" owner="Academia Sinica"
   version="1.6" >
    <LexicalEntry>
      <Lemma writtenForm="自然" partOfSpeech="n">
      </Lemma>
      <Sense id="自然_1" synset="zho-16-03059301-n">
      </Sense>
    </LexicalEntry>
    ……
    <Synset id="zho-16-03059301-n" baseConcept="3">
      <Definition gloss="普通名詞。天然生成的環境與事物。">
        <Statement example="風景畫家走出工作室，開始描繪
        戶外的自然" />
      </Definition>
      <SynsetRelations>
        <SynsetRelation target="zho-16-06653061-n"
        relType="has_synonym"> (大自然_1)
      </SynsetRelation>
      <MonolingualExternalRefs>
        <MonolingualExternalRef externalSystem="SUMO"
        externalReference="(ComplementFn)InternationalProcess"
        relType="plus" />
      </MonolingualExternalRefs>
    </Synset>
    ……
  </Lexion>
  <SenseAxes>
    <SenseAxis id="sa_zho16-eng30_17638" relType="eq_synonym">
      <Target ID="zho-16-03059301-n"> (自然_1)
      <Target ID="eng-30-11408559-n"> (nature)
    </SenseAxis>
    ……
  </SenseAxes>
</LexicalResource>
```

Figure 3: The lemma 自然 in CWN-LMF format.

## 5 Quantitative Analysis of CWN-LMF

There are 8,646 lemmas / 25,961 senses until December 2008 have been analyzed in CWN 1.6. So far the work on Chinese word distinction is still ongoing. It is expected that there are more analyzed results in the next released version.

Among analyzed 25,961 senses, there are 268 senses and 1,217 senses that belong to the first-class and the second –class basic words, respectively. When part-of-speech is concerned, we can find most of these senses belong to nouns or verbs. There are 12,106 nouns, 10,454 nouns, 806 adjectives and 1,605 adverbs in CWN 1.6

We further distinguish semantic relations of CWN 1.6 and found that there are 3,328 synonyms, 213 near synonyms, 246 hypernyms, 38 hyponyms, 3 holonyms, 240 paranyms, 369 antonyms and 432 variants, respectively.

The IEEE SUMO is the only external system for monolingual references in CWN-LMF. There are 21,925 senses that were pointed to SUMO so far. In addition, there are 17,952 senses which shared the same equivalent relation to Princeton WordNet 3.0 in CWN-LMF.

## 6 Application Scenarios

The EU-7 project, KYOTO (Knowledge Yielding Ontologies for Transition-based Organization), wants to make knowledge sharable between communities of people, culture, language and computers, by assigning meaning to text and giving text to meaning (Vossen et al., 2008a; 2008b). The goal of KYOTO is a system that allows people in communities to define the meaning of their words and terms in a shared Wiki platform so that it becomes anchored across languages and cultures but also so that a computer can use this knowledge to detect knowledge and facts in text.

KYOTO is a generic system offering knowledge transition and information across different target groups, transgressing linguistic, cultural and geographic boundaries. Initially developed for the environmental domain, KYOTO will be usable in any knowledge domain for mining, organizing, and distributing information on a global scale in both European and non-European languages.

Whereas the current Wikipedia uses free text to share knowledge, KYOTO will represent this knowledge so that a computer can understand it. For example, the notion of environmental *footprint* will become defined in the same way in all these languages but also in such a way that the computer knows what information is necessary to calculate a *footprint*. With these definitions it will be possible to find information on footprints in documents, websites and reports so that users can directly ask the computer for actual information in their environment, for instance, what is the footprint of their town, their region or their company.

KYOTO's principal components are an ontology linked to WordNets in seven different languages (Basque, Chinese, Dutch, English, Italian, Japanese and Spanish). Due to different natures of languages, the different designed architectures were used to develop WordNets in theses languages. A unified framework is needed for information exchange. LMF is hence adopted as the framework at lexical semantic level in this project. The WordNet in these languages are compiled with designed WordNet-LMF format. CWN-LMF will also be involved and benefit for cross-language interpretabilities in semantic search field.

## 7 Discussion

Due to characters of Chinese language, there are some difficulties of encoding Chinese WordNet into Wordnet-LMF. A brief description is presented as follows.

Chinese WordNet was designed for Chinese word sense distinction and its lexical semantic relationships. The designed architecture belongs to word-driven, not synset-driven. So in CWN-LMF, we encoded a sense as an individual synset and marked up the "has_synonym" relation when senses belong to the same WordNet synset.

In addition, how to define the basic concept of Chinese language is difficult. So far the basic word lists of the NEDO project were used as preliminary basis. We need a further method to distinguish *baseConcept* attribute of word senses.

## 8 Conclusions

This study describes the design and implementation of how the Wordnet-LMF used to represent lexical semantics in Chinese WordNet. CWN-LMF is benefit for data exchange among computational linguistic resources, and also promises a convenient uniformity for domain-specific applications such as KYOTO in cross-language semantic search field.

Future work is investigated with several directions. We are planning to release Chinese WordNet 1.6 using CWN-LMF format in an xml file, including a XML DTD in the following days. In addition, the use of this lingual resource for further linguistic research is also under investigation.

## Acknowledgements

The authors would like to thank Prof. Claudia Soria for her constructive comments. This work was funded by National Science Council, Taiwan under Grants NSC 97-2923-I-001-001-MY3., and also cooperated with EU-FP7 KYOTO project.

## References

CKIP. 2003. Sense and Sensibility Vol. I. Technical Report 03-01. Taipei: Academia Sinica.

Fellbaum, C.. 1998. WordNet: an Electronic Lexical Database. The MIT Press.

Francopoulo, G., Bel, N., George, M., Calzolari, N., Monachini, M., Pet, M. and Soria, C.. 2006a. Lexical Markup Framework (LMF) for NLP Multilingual Resources. Proceedings of COLING-ACL Workshop on Multilingual Language Resources and Interoperability.

Francopoulo, G., Bel, N., George, M., Calzolari, N., Monachini, M., Pet, M. and Soria, C.. 2006b. LMF for Multilingual, Specialized Lexicons. Proceedings of the LREC Workshop on Acquiring and Representing Multilingual, Specialized Lexicons: the Case of Biomedicine.

Francopoulo, G., Bel, N., George, M., Calzolari, N., Monachini, M., Pet, M. and Soria, C.. 2009. Multilingual Resources for NLP in the Lexical Markup Framework (LMF). Language Resource and Evaluation. 43:57-70.

Huang, C.-R., Chang, R.-Y. and Lee, H.-P.. 2004. Sinica BOW (Bilingual Ontological Wordnet): Integration of Bilingual WordNet and SUMO. Proceedings of the 4th International Conference on Language Resources and Evaluation.

Huang, C.-R., Chen, C.-L., Weng, C.-X., Lee, H.-P., Chen, Y.-X. and Chen, K.-J.. 2005. The Sinica Sense Management System: Design and Implementation. Computational Linguistics and Chinese Language Processing. 10(4): 417-430.

Huang, C.-R., Hsieh, S.-K., Hong, J.-F., Chen, Y.-Z., Su, I.-L., Chen, Y.-X. and Huang, S.-W.. 2008. Chinese Wordnet: Design, Implementation, and Application of an Infrastructure for Cross-lingual Knowledge Processing. Proceedings of the 9th Chinese Lexical Semantics Workshop.

Huang, C.-R., Lee, H.-P. and Hong, J.-F.. 2004. Domain Lexico-Taxonomy: an Approach Towards Multi-domain Language Processing. Proceedings of the Asian Symposium on Natural Language Processing to Overcome Language Barriers.

Huang, C.-R., Lee, H.-P. and Hong, J.-F.. 2005. The Robustness of Domain Lexico-Taxonomy: Expanding Domain Lexicon with Cilin. Proceedings of the 4th ACL SIGHAN Workshop on Chinese Language Processing.

Huang, C.-R., Su, I.-L., Hsiao, P.-Y. and Ke, X.-L.. 2008. Paranymy: Enriching Ontological Knowledge in Wordnets. Proceedings of the 4th Global WordNet Conference.

Huang, C.-R., Tsai, D. B.-S., Weng, C.-X., Chu, N.-X., Ho, W.-R., Huang, L.-H. and Tsai, I.-N.. 2003. Sense and Meaning Facet: Criteria and Operational Guidelines for Chinese Sense Distinction. Proceedings of the 4th Chinese Lexical Semantics Workshop.

LMF. 2009. Lexical Markup Framework. ISO-24613. Geneva:ISO.

Magnini, B. and Cavaglia, G.. 2000. Integrating Subject Field Codes into WordNet. Proceedings of the 2nd International Conference on Language Resources and Evaluation.

Miller, G. A.. 1995. WordNet: a Lexical Database for English. Communications of the ACM. 38(11): 39-41.

Niles, I. and Pease, A.. 2001. Toward a Standard Upper Ontology. Proceedings of the 2nd International Conference on Formal Ontology in Information Systems.

Soria, C., Monachini, M. and Vossen, P.. 2009. Wordnet-LMF: Fleshing out a Standardized Format for Wordnet Interoperability. Proceedings of ACM Workshop on Intercultural Collaboration.

Soria, C., Monachini, M., Bertagna, F., Calzolari, N., Huang, C.-R., Hsieh, S.-K., Marchetti, A. and Tesconi, M.. 2009. Exploring Interoperability of Language Resources: the Case of Cross-lingual Semi-automatic Enrichment of Wordnets. Language Resource and Evaluation. 43:87-96.

Strapparava, C. and Valitutti, A.. 2004. WordNet-Affect: an Affective Extension of WordNet. Proceedings of the 4th International Conference on Language Resources and Evaluation.

Tokuaga, T., Sornlertlamvanich, V., Charoenporn, T., Calzolari, N., Monachini, M., Soria, C., Huang, C.-R., Yu, Y., Yu, H. and Prevot, L.. 2006. Infrastructure for Standardization of Asian Language Resources. Proceedings of the COLING/ACL Main Conference Poster Sessions.

Vossen, P.. 2004. EuroWordNet: a Multilingual Database of Autonomous and Language-specific Wordnets Connected via an Inter-Lingual-Index. International Journal of Linguistics. 17(2): 1-23.

Vossen, P., Agirre, E., Calzolari, N., Fellbaum, C., Hsieh, S.-K., Huang, C.-R., Isahara, H., Kanzaki, K., Marchetti, A., Monachini, M., Neri, F., Raffaelli, R., Rigau, G., Tescon, M. and VanGent, J.. 2008a. KYOTO: A System for Mining, Structur-

ing, and Distributing Knowledge Across Languages and Cultures. Proceedings of 6th International Conference on Language Resource and Evaluation.

Vossen, P., Agirre, E., Calzolari, N., Fellbaum, C., Hsieh, S.-K., Huang, C.-R., Isahara, H., Kanzaki, K., Marchetti, A., Monachini, M., Neri, F., Raffaelli, R., Rigau, G., Tescon, M. and VanGent, J.. 2008b. KYOTO: A System for Mining, Structuring, and Distributing Knowledge Across Languages and Cultures. Proceedings of the 4th International Global WordNet Conference.