Chinese FrameNet and OWL Representation

Liping You[1], Tao Liu[2], Kaiying Liu[3]
[1, 2] *College of Management, Shanxi University, Taiyuan, 030006*
[3] *School of Computer and Information Technology, Shanxi University, Taiyuan, 030006*
Email: yoliping@163.com

Abstract

The Chinese FrameNet Project is creating an lexical resource for Chinese, based on the principles of Frame Semantics and supported by corpus evidence. Our description of each lexical item identifies the frames which underlie a given meaning and the ways in which the Frame Elements are realized in structures headed by the word, through manual annotation of example sentences and automatic summarization of the resulting annotations. It will be available in XML, and can be displayed and queried via the web and other interfaces. We are trying to translate the CFN data into the OWL Web Ontology Language, a revision of the DAML+OIL web ontology language, which can represent our ontologies and to make CFN information machine readable and understandable. This paper reports on our representation of this data in OWL.

1. Introduction

The development of lexical semantic resources is widely recognized as a prerequisite to the progress in natural language understanding. Sophisticated efforts in this direction include WordNet [1], MindNet [2] and FrameNet [3], all of which are English lexical semantic resources. In China, some researchers put effort into building Chinese lexical semantic resources: the Chinese Concept Dictionary (CCD) [4], developed by Peking University, Beijing, is a Chinese-English bilingual concept dictionary with the framework of WordNet and HowNet [5], developed by Zhendong Dong, is an on-line common-sense knowledge base unveiling inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents. However, it's a long way for Chinese lexical resources to be sufficient for application.

The Chinese FrameNet Project [6] is creating an lexical resource for Chinese, based on the principles of Frame Semantics and supported by corpus evidence. The CFN database now contains information about more than 1760 lexical units (senses of words) in more than 130 frames, based on the annotation of more than 10,000 sentences. This paper will explain the theory behind CFN, briefly discuss the annotation process, and then describe how the CFN data can be represented in OWL, so that it will become a resource for the Semantic Web.

2. Chinese FrameNet

2.1 Background: Frame Semantics

The basic assumption of Frame Semantics [7][8] as it applies to the description of lexical meanings is that each word (in a given meaning) evokes a particular frame and possibly profiles some element or aspect of that frame. An “evoked” frame is the structure of knowledge required for the understanding of a given lexical or phrasal item; a “profiled” entity is the component of a frame that integrates directly into the semantic structure of the surrounding text or sentence. The frames in question can be simple—small static scenes or states of affairs, simple patterns of contrast, relations between entities and the roles they serve, or possibly quite complex event types—which we can call scenarios—that provide the background for words that profile one or more of their phases or participants.

The notion of frame can be exemplified with the Commercial Transaction Frame, whose elements, Frame Elements (FEs), include a buyer, a seller, goods, and money. (Note that these FEs have been designated in terms of situational roles; this contrasts with the notion of semantic roles as articulated in case grammar.) Among the large set of semantically related verbs linked to this frame are *买(buy), 卖(sell), 支付(pay), 花费(spend),* and *收款(charge),* each of which indexes or evokes different aspects of the frame. The verb *买(buy)* focuses on the buyer and the goods,

0-7695-2930-5/07 $25.00 © 2007 IEEE
DOI 10.1109/ALPIT.2007.50

backgrounding the seller and the money; *卖(sell)* focuses on the seller and the goods, backgrounding the buyer and the money; *支付(pay)* focuses on the buyer, the money, and the seller, backgrounding the goods; and so on. The idea is that knowing the meaning of any one of these verbs requires knowing what takes place in a commercial transaction and knowing the meaning of any one verb means, in some sense, knowing the meaning of all of them. The knowledge and experience structured by the Commercial Transaction Frame provide the background and motivation for the categories represented by the words. The words, that is, the linguistic material, **evoke** the frame (in the mind of a speaker/hearer); the interpreter (of an utterance or a text in which the words occur) **invokes** the frame.

A complete description of these verbs must also include information about their grammatical properties and the various syntactic patterns in which they occur. What elements or aspects of the frame may be realized as the subject of the verb, as its object, if there is one, and what will be the surface form of the others? Which of these elements are optional and which are obligatory? For example, in the sentence *李丽用100块钱从张阿姨那买了一条狗 (Lili bought a dog from aunty Zhang for RMB 100),* the subject, *李丽 (Lili)*, is the buyer and object, *一条狗 (a dog)*, is the goods; both elements are obligatory. The backgrounded elements, both of which are optional, surface as adverbial modifiers: *用100块钱 (for RMB100)*, the money, and *从张阿姨那 (from aunty Zhang)*, the seller. Of note are the different prepositions used with each, particularly that *从 (from)* is the preposition allowing the interpretation that *张阿姨 (aunty Zhang)* is the seller, while other prepositions may be used for the money--e.g. *用, 以 (for).* Such grammatical information about the syntactic-semantic valence description of each verb is not specified in the frame. Nevertheless, it is deducible from the rich descriptions of the different FEs.

The job of CFN is to produce frame-semantic descriptions of several thousand Chinese lexical items and to document from attested instances of contemporary Chinese the manner in which FEs are grammatically instantiated in Chinese sentences and to organize and exhibit the results of such findings in a systematic way. The corpus used in CFN so far has been the CCL Corpus (corpus of the Center for Chinese Linguistics PKU [9]), more than 115,000,000 characters of contemporary Chinese.

The CFN database can be seen both as a dictionary and a thesaurus. It is a dictionary in that each lexical unit is provided with (1) the name of its frame, (2) a definition, (3) a valence description which summarizes the attested combinatorial possibilities, and (4) access to annotated examples illustrating each syntactic pattern found in the corpus and the kinds of semantic information instanced with such patterns. It is a thesaurus in that, by being linked to frames, each word is directly connected with other words in its frame(s), and further extensions are provided by working out the ways in which a word's basic frames are connected with other frames through relations of inheritance (possibly multiple inheritance) and composition.

2.2 An Example Frame: 自主感知 (Perception_active)

For instance, to understand the Chinese verbs below, one will evoke the same frame, 自主感知 (Perception_active), in his mind.

听 v (to listen), 看 v (to look), 尝 v (to taste), 闻 v (to smell), 嗅 v (to smell), 听听 v (to have a listen), 闻闻 v (to have a smell), 看看 v (to have a look), 尝尝 v (to have a taste), 眺望 (to overlook), 观看 v (to watch), 偷看 v (to peep), 窥视 v (to peep), 偷窥 v (to peep), 欣赏 v (to watch), 观赏 v (to watch), 偷听 v (to eavesdrop), 感觉 v (to perceive), 感受 v (to perceive), 凝视 v (to gaze), 一瞥 v (to have a glance), 观察 v (to observe), 探察 v (to scout), 品尝 v (to taste), 侦察 v (to scout), 斜视 v (to look sideways), 盯 v (to stare), 瞅 v (to watch), 观 v (to watch)

All the above words involve events that perceivers intentionally direct their attention to some entity or phenomenon in order to have a perceptual experience. So they could be categorized in one frame, called 自主感知 (Perception_active). The FEs are defined like Table 1.

Core FE is one that instantiates a conceptually necessary component of a frame, while making the frame unique and different from other frames. FEs that do not introduce additional, independent or distinct events from the main reported event are characterized as non-core FEs. Non-core FEs mark such notions as Time, Place, Manner, Means, Degree, etc..

Table 1. The FEs of the Perception_active frame

Core	
身体部位 (Body_part)	Body_part identifies the part of the body used as a sensory organ of the Perceiver_agentive.
方向 (Direction)	Direction is used for all path-like expressions that describe how the perceiver's attention is directed during the act of perception.
自主感知者 (Perceiver_agentive)	The Perceiver_agentive performs some action in order to have a perceptual experience.
现象 (Phenomenon)	Phenomenon indicates the entity or phenomenon to which the Perceiver_agentive directs his or her attention in order to have a perceptual experience.
Non-core	
形容 (Depictive)	Depictive is used for predicate expressions that apply to the Phenomenon, providing some information about the state it is in while the perceiver's attention is directed to it:
动作时间量 (Duration _of _action)	This FE identifies the Duration of time for which the Perception takes place.
背景 (Ground)	Ground is the perceptual background against which the Phenomenon is experienced by the Perceiver_agentive.
感知者位置 (Location_of_perceiver)	This FE is the position of the Perceiver during the act of perception.
...	...

2.3 Relations between Frames and FEs

Inheritance Frame inheritance is an IS-A relation. If frame B inherits from frame A, then B elaborates A, and is a subtype of A. Furthermore, all the FEs of A are inherited by B. They may appear in B with exactly the same names and semantic types, or they may be subtypes (elaborations) of the FEs in A. The picture is further complicated by situations of multiple inheritance; an FE in one frame may inherit from FEs in two parent frames, combining their semantics. For example, the sentence *科学家们正对核武器进行观察 (The scientists are observing nuclear weapons)* belongs to the 自主感知 (Perception_active) frame, which inherits from both 感知觉 (Perception) and 有意行为 (Intentionally_act). Thus, *科学家们 (the scientists)* is both a Perceiver by virtue of inheritance from Perception and a Agent by virtue of inheritance from Intentionally_act. Fig 1 shows some inheritance relation of the Perception_active frame.

Subframes The other type of relation between frames which is currently represented in the CFN database is between a complex frame and several simpler frames (subframes) which constitute it. We call this frame composition, by which we mean the situation in which a complex frame is made up of parts which are also frames. In such cases, FEs of the complex frame may be identified (mapped) to the FEs of the subparts, although not all FEs of one need have any relation to the other. (In this respect, it contrasts with inheritance.) For example, in a perceiving process (the Perception frame), first, something acts on one’s sense organ with sound or light (the Make_noise frame

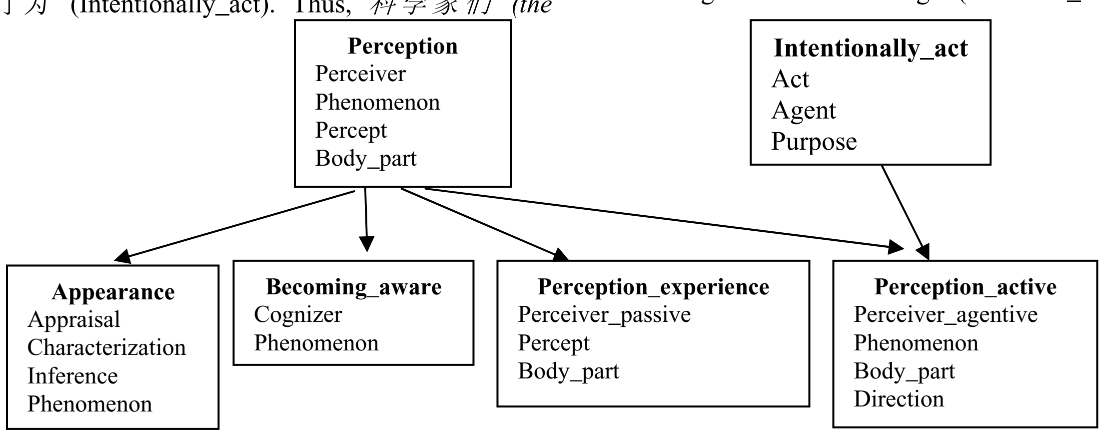

Figure 1. The Perception_active frame and some of its inherited frames.

and the Location_of_light frame). Then he becomes aware it (the Becoming_aware frame) and differentiate it from other entities in his memory (the Differentiation frame). At last, he construes it as belonging to a certain Category (Categorization). Thus, the Perception frame has subframes of Location_of_light, Make_noise, Becoming_aware, Differentiation and Categorization.

2.4 Annotated sentences

Sentences are marked up to exemplify the semantic and syntactic properties of the lexical items. For example, when given the target verb *眺望 (to overlook)*, which belongs to the Perception_active frame, we select the sentence *当我夜间从阳台上眺望满城灿烂的灯火时，我仿佛又看见了她的波光 (When I overlooked the myriad twinkling light in the city at night from the balcony, it seemed as if I saw her eyes),* and mark the FEs and the phrase types and grammatical functions of the constituents instantiating them. The resulting annotations can be displayed in a bracketed notation like this:

当<Perceiver_agentive-np-subj 我> <Time-tp-adva 夜间> <Location_of_perceiver-pp-adva 从阳台上> <tgt 眺望> < Phenomenon-np-obj 满城灿烂的灯火> 时，我仿佛又看见了她的波光。

3. An OWL representation of CFN Frames

The World Wide Web (WWW) contains a large amount of information which is expanding at a rapid rate. Most of that information is currently being represented using the Hypertext Markup Language (HTML), which is designed to allow web developers to display information in a way that is accessible to humans for viewing via web browsers. While HTML allows us to visualize the information on the web, it doesn't provide much capability to describe the information in ways that facilitate the use of software programs to find or interpret it. The World Wide Web Consortium (W3C) has developed the Extensible Markup Language (XML) which allows information to be more accurately described using tags. The use of XML to provide metadata markup makes the meaning of the word unambiguous. However, XML has a limited capability to describe the relationships (schemas or ontologies) with respect to objects. The use of ontologies provides a very powerful way to describe objects and their relationships to other objects. OWL Web Ontology Language can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. OWL has more facilities for expressing meaning and semantics than XML, RDF, and RDF-S, and thus OWL goes beyond these languages in its ability to represent machine interpretable content on the Web. OWL is a revision of the DAML+OIL web ontology language. We are trying to translate the CFN data to semantic web language, so that it could be machine readable and understandable. We expect it to become useful for various applications on the Semantic Web.

OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full. *OWL Lite* supports those users primarily needing a classification hierarchy and simple constraints. *OWL DL* supports those users who want the maximum expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time). *OWL Full* is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees. Each of these sublanguages is an extension of its simpler predecessor, both in what can be legally expressed and in what can be validly concluded.

Our Frame declaration begins with an RDF start tag including several namespace declarations of the form:

```
1 <?xml version="1.0" encoding="gb2312"?>
2 <!DOCTYPE rdf:RDF[<!ENTITY xsd
  "http://www.w3c.org/2001/XMLSchema#">
3   <!ENTITY rdfs "www.w3c.org/2000/01/rdf-schema"
>
4   <!ENTITY dc "http://purl.org/dc/elements/1.1/">]>
5 <rdf:RDF xmlns   ="http://www.CFN.org/Frame/"
6   xmlns:frame="http://www.CFN.org/Frame/"
7   xml:base="http://www.CFN.org/Frame/"
8   xmlns:words="http://www.CFN.org/Frame/words"
9   xmlns:owl="http://www.w3.org/2002/07/owl#"
10  xmlns:rdf="www.w3c.org/199/02/22-rdf-syntax-ns#"
11  xmlns:rdfs= "http://www.w3.org/2000/01/rdf-
schema#">
```

Line 1 is the XML version declarations of this document, line 2-4 describe the document's type, which is the namespace of xsd, rdf and dc， and line 5-11 declare the name space of the terms in this document.

Then we define the URI references as below:

```
12 <owl:Ontology rdf:about="">
13   <rdfs:comment>Frame Ontology 0f
       CFN</rdfs:comment>
14   <owl:imports
       rdf:resource="http://www.CFN.org/Frame/words"/>
15   <rdfs:label>Frame Ontology</rdfs:label>
16 </owl:Ontology>
```

Line 14 makes this document refer to some terms in the lexical database. There are shared lexical units in both of the frame database and the lexical database, so they can refer to each other.

The most general object of interest is a frame. We define the FRAME class as a owl:unionOf a BACKGROUNDFRAME class (such as Perception frame) and a LEXICALFRAME class (such as the Perception_active frame). We then define a bunch of bookkeeping subclasses, properties and restrictions on the FRAME class. An example of the name property is shown below.

```
<owl:class rdf:ID="frame">
  <rdfs:comment>the most general class</rdfs:comment>
  <owl:unionof rdf:parsetype="collection">
   <owl:class rdf:about="backgroundframe"/>
   <owl:class rdf:about="lexicalframe"/>
  </owl:unionof>
  <rdfs:subclassof>
  <owl:Restriction>
   <owl:onProperty rdf:resource="#hasname" />
   <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCar
dinality>
  </owl:Restriction>
……
</owl:class>
```

We define the FE class as a subclass of the frame and it has its own subclasses: core FE and non-core FE, as well as their restrictions.

```
<owl:class rdf:ID="FE">
 <rdfs:subclassof rdf:ID="#frame"/>
 <rdfs:subclassof>
   <owl:Restriction>
     <owl:onProperty rdf:resource="#hasfullname"/>
     <owl:cardinality
    rdf:datatype="&xsd;nonNegativeInteger">0</owl:ca
    rdinality>
   </owl:Restriction>
 </rdfs:subclassof>
 <rdfs:subclassof>
   <owl:Restriction>
     <owl:onProperty rdf:resource="#hasabbrname"/>
     <owl:cardinality
    rdf:datatype="&xsd;nonNegativeInteger">1</owl:ca
    rdinality>
   </owl:Restriction>
 </rdfs:subclassof>
 <rdfs:subclassof>
   <owl:restriction>
     <owl:objectProperty rdf:ID="hascoreelement"/>
     <owl:minCardinality
    rdf:datatype="&xsd;nonNegativeInteger">1</owl:m
    inCardinality>
   </owl:restriction>
 </rdfs:subclassof>
 <rdfs:subclassof>
   <owl:Restriction>
   <owl:objectProperty
rdf:resource="hasnoncoreelement">
     <rdfs:range rdf:resource="&rdfs;literal"/>
   </owl:objectProperty>
     <owl:mincardinality
     rdf:datatype="&xsd;nonNegativeInteger">0</owl:min
     cardinality>
     </owl:Restriction>
  </rdfs:subclassof>
 </owl:class>
```

In CFN, a frame may inherit (A ISA B) from other frames or be a composition of a set of subframes (which are frames themselves). We define these relations as a subclass of the frame and use the OWL DL to define the specific relation.

```
<owl:class rdf:ID="relation">
  <rdfs:subclassof rdf:resource="#frame"/>
  <owl:unionof rdf:parsetype="collection">
   <owl:class rdf:about="#inheritance"/>
   <owl:class rdf:about="#subframe"/>
  </owl:unionof>
</owl:class>
```

After the structure description above, the specific frame instances would be represented. In the example fragment below we show the OWL definition of a frame name: 自主感知 (Perception_active)

```
<owl:Class rdf:ID="framename">
   <rdfs:comment>框架的英文名称</rdfs:comment>
   <owl:oneOf rdf:parseType="Collection">
  <owl:Thing rdf:resource="#自主感知"/>
   </owl:oneOf>
</owl:Class>
```

In the example fragment below we show part of the definition about the specific FEs and their descriptions, including the FE's Chinese name and its English one and abbriviate one. Limited to the scope of this paper, here we only show some of the FEs representation of the Perception_active frame.

```
<FE rdf:ID="自主感知者">
   <hasfullname rdf:resourse="#Perceiver_agentive"/>
   <hasabbrname rdf:resourse="#Perc_Agt"/>
</FE>
   <FE rdf:ID="现象">
   <hasfullname rdf:resourse="#Phenomenon"/>
   <hasabbrname rdf:resourse="#Phen"/>
</FE>
```

Below is the instance description about the specific frame Perception_active.

```
<frame rdf:ID="框架 1">
  <hasname rdf:resourse="#自主感知"/>
  <hascreator rdf:resourse="#由丽萍"/>
  <hasdate rdf:resourse="#FEB_20_2006"/>
  <hasFE rdf:resourse="#自主感知框架元素"/>
  <hasRelation rdf:resourse="#自主感知框架关系"/>
</frame>
<FE rdf:ID="自主感知框架元素">
  <hascoreelelment rdf:resourse="#自主感知者"/>
  <hascoreelelment rdf:resourse="#现象"/>
…….
</FE>
```

```
<Relation rdf:ID="自主感知框架关系">
  <hasparent rdf:resource="#有意行为"/>
  <hasparent rdf:resource="#感知觉"/>
</Relation>
```

At last, the lexical units are represented in the example fragment below. All of them can be linked to another type of OWL document which specifically represents the lexical units' information. Thus we use # to represent this link.

```
<words:lexical rdf:resource="#听"/>
<words:lexical rdf:resource="#看"/>
……
<words:lexical rdf:resource="#瞅"/>
<words:lexical rdf:resource="#观"/>
```

4. Conclusion

The World Wide Web (WWW) contains a large amount of information which is expanding at a rapid rate. The information contained encompasses diverse ranges and types of data, from structured databases to text. While XML allows for the specification of metadata as a means to specify the structure and syntax of a URL, it does not provide much capability to describe the semantic relations between the different pieces of information in a manner that can be exploited by software programs for automation or interpretation. OWL is a widely used language related to the Semantic Web initiative that is poised to remedy this situation. This rapid growth of the web also engenders need for sophisticated techniques to represent lexical and sense distinctions in a machine readable and interpretable manner. The CFN database documents the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses, through manual annotation of example sentences and automatic summarization of the resulting annotations. We believe that CFN offers promise as a potential resource to aid in the automatic identification and disambiguation of Chinese word meanings on the semantic web. This paper described an encoding of CFN data in the OWL language. We expect the OWL encoding of CFN frames to become useful for various applications on the semantic web.

Acknowledgements

This work is supported in part by the National 863 Plan (2006AA01Z142) and the Shanxi University research fund (0609099).

References

[1] http://wordnet.princeton.edu/
[2] http://research.microsoft.com/nlp/Projects/MindNet.aspx
[3] http://framenet.icsi.berkeley.edu/
[4] Yu Jiangsheng, Yu Shiwen, "The Structure of Chinese Concept dictionary", *Journal of Chinese Information Processing*, Beijing, 2002, 16 (4). pp. 12-20.
[5] http://www.keenage.com/
[6] Liping YOU, Kaiying LIU, "Building Chinese FrameNet Database", *in proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE)*, 2005, pp. 301-306.
[7] Fillmore Charles J., "Frame Semantics and the Nature of Language", *Origins and Evolution of Language and Speech*, Annals of the NY Academy of Sciences, Vol. 280.
[8] Fillmore, C.J., "Frame Semantics". *The Linguistic Society of Korea*, Linguistics in the Morning Calm. Seoul: Hanshin. 1982.
[9] http://ccl.pku.edu.cn:8080/ccl_corpus/jsearch/index.jsp?dir=xiandai
[10] Deborah L. McGuinness Frank van Harmelen, OWL Web Ontology Language Overview, http://www.w3.org/TR/2004/REC-owl-features-20040210/, 2004,02(10).