UNL2005
from Language Infrastructure toward Knowledge Infrastructure

Hiroshi Uchida
UNL Center
UNDL Foundation
53-70, Jingumae 5-chome
Shibuya-ku, Tokyo 150-8925
uchida@undl.org

Meiying Zhu
UNL Center
UNDL Foundation
53-70, Jingumae 5-chome
Shibuya-ku, Tokyo 150-8925
zhu@undl.org

1. Introduction

The UNL was born within the United Nations and was conceived at the Institute of the Advanced Studies of the United Nations University(UNU/IAS). The UNL have been developed under the UNL Programme started in 1996, as an initiative of the UNU/IAS.

UNL is an acronym for “Universal Networking Language”. The UNL is a computer language that enables computers to process information and knowledge across the language barriers. It is an artificial language that replicates the functions of natural languages in human communication. As a result, it enables people to express all knowledge conveyed by natural languages. It also enables computer to intercommunicate, thus providing people with a Language Infrastructure (LI) for distributing, receiving and understanding multilingual information. Such multilingual information can be accessed by natural languages through the UNL System.

UNL is a language for representing information and knowledge provided by natural languages. It has all the components corresponding to that of a natural language. It is composed of words expressing concepts called "Universal words", also referred to as UWs that are inter-linked with other UWs to form the UNL expressions of sentences. These links, called as "relations", specify the role of each word in a sentence. The subjective meanings intended by the author are expressed through "attributes".

When computers make intelligent processing based on knowledge including reasoning, enough knowledge must be provided in the form that accessible and understandable for computers. We call the collection of such knowledge “Knowledge Infrastructure (KI)”. UNL 2005 is improved for meeting this need that UNL must be able to represent knowledge for providing KI for intelligent processing.

Major enhancements of UNL2005 from Language Infrastructure toward Knowledge Infrastructure are the following:

- Introducing the function of describing inter-sentence relations and cross-reference of UWs, sentences and paragraphs for expressing the total meaning of UNL expressions;
- Introducing knowledge representation structure; and
- Introducing the function of describing Logical Expression in UNL

2. How UNL Express Information and Knowledge

UNL expresses information and knowledge in the form of semantic network with hyper-node. Differently from natural languages, UNL expressions are unambiguous. In the UNL semantic network, nodes represent concepts and arcs represent relations between concepts. Concepts can be annotated. Such semantic network is called "UNL expression".

The unit of UNL expression is a UNL document. A UNL document is considered as a hyper-node composed of a semantic network among sentences or paragraphs. A paragraph and a sentence are hyper-nodes too. A hyper-node of a paragraph is composed of a semantic network among sentences or other paragraphs. A hyper-node of a sentence is composed of a semantic network among Universal Words (UWs).

The semantic network of the hyper-node of a UNL document consists of two parts: a sequence of hyper-nodes of paragraphs or sentences of the document, and a set of semantic relations among hyper-nodes of the paragraphs or sentences. A sequence of hyper-nodes can be considered as a directed (ordered) graph linked with the relation "nxt" which express the physical sequence of paragraphs or sentences. Likewise, the semantic network of a hyper-node of a paragraph also consists of two parts: a sequence of hyper-nodes of sentences or other paragraphs included in the paragraph, and a set of semantic relations among hyper-nodes of the paragraphs or sentences. The semantic network of a hyper-node of a sentence consists of a set of semantic relations between UWs.

Every hyper-node of a paragraph or sentence and UW can be referred to from any other hyper-node of a paragraph or a sentence or a UW.

2.1. UNL Expression

UNL expression is a semantic network made up of a set of **binary relations**, each binary relation is composed of a relation and two UWs that hold the relation. A binary relation of UNL is expressed in the following way:

<relation> (<uw1>, <uw2>)

In <relation>, one of the relations defined in the UNL Specifications is described. In <uw1> and <uw2>, the two UWs that have the relation given by <relation> are described. A semantic network of UNL expression is a **directed graph** composed of binary relations with direction. The three elements of a binary relation have the following interrelationship:

<uw1> -- <relation> -> <uw2>

This binary relation is **interpreted** as that: the UW given in <uw2> plays the role indicated by the relation given in <relation> held by the UW given in <uw1>; whereas the UW given in <uw1> holds the relation given in <relation> with the UW given in <uw2>.

Hyper-nodes are allowed in the semantic network of UNL expression. That is, each node in a graph, <uw1> and <uw2> of a binary relation, can be hyper node containing a semantic network. Such a hyper node made up of a semantic network of UNL expression is called a "**scope**". A scope can be connected with other UWs or scopes because a scope is considered as a UW. A binary relation in a scope is distinguished from others by assigning an ID to the <relations>.

The **general format** of binary relations of UNL expression allowing scopes is the following:

<relation : <scope-ID> (<node1> , <node2>)

A UNL expression of a sentence is identified with the following tags: {unl} and {/unl}. Any component, such as a word, phrase and, of course, a sentence of a natural language can be represented as UNL expressions. A UNL expression therefore consists of a UW or a (set of) binary relation(s). In UNL documents, a UNL expression for a sentence is enclosed by the tags {unl} and {/unl} inside [S] and [/S]. If a UNL expression consists of a UW, this UW should be enclosed further by the tags [W] and [/W]. If necessary, the whole sentence can also be expressed as a scope. In this case, the Scope-ID of the scope should be enclosed by [W] and [/W].

A UNL expression consists of a set of binary relations	{unl} <binary relation> ... {/unl}
A UNL expression consists of a UW	{unl} [W] <UW><attribute list> [/W] {/unl}
A UNL expression consists of a scope	{unl} [W] ": "<Scope-ID><attribute list> [/W] <binary relation> ... {/unl}

Descriptive format of a binary relation is the following:

<binary relation> ::= <relation> [":"<Scope-ID>] "("
{{ <UW$_1$> [":" <UW-ID$_1$>]} | { ":" <Scope-ID$_1$> }}
[<attribute list>] ","{{ <UW$_2$> [":" <UW-ID$_2$>]} |
{ ":" <Scope-ID$_2$> }}[<attribute list>] ")"

or

::= <relation> [":"<Scope-ID>] "("
{{ <UW$_1$> [":" <UW-ID$_1$>]} | { ":" <Scope-ID$_1$> }}
[<attribute list>] ","<referee node> ")"

or

::= <relation> [":"<Scope-ID>] "("<referee node> ","
{{ <UW$_2$> [":" <UW-ID$_2$>]} | { ":" <Scope-ID$_2$> }}
[<attribute list>] ")"

A UNL expression can include more than one scope. **Scope-IDs** are for identifying each concept specified by scopes in a UNL expression. A scope is a group of binary relations that can be referred to as a UW by indicating its Scope-ID.

UW-IDs are for identifying each concept specified by UWs in a UNL expression. If a UW appears in a UNL expression more than once and means different concepts (things or events), a unique UW-ID must be given to each concept of the UWs.

2.2. UNL Document

UNL expressions are described in UNL documents. A UNL document is a text file that includes the original sentences, UNL expressions, sentences in target languages, and tags of UNL document.

A UNL document is enclosed with tags “[D:<dinf>]” and “[/D]”. Within these tags, each paragraph is enclosed with a pair of tags “[P:<p_num>]” and “[/P]”, and each sentence is enclosed with a pair of tags “[S:<s_num>]” and “[/S]”. Inside a sentence, the text of original sentence is enclosed with “{org:<l_tag>}” and “{/org}”, its UNL expression is enclosed with “{unl:<uinf>}” and “{/unl}”. Sentences of target languages can also be stored in the UNL document. Each target sentence is enclosed with a pair of language tags “{<l_tag>}” and “{</l_tag>}” following the UNL expression of each sentence.

```
<UNL Document>      ::= "[D:" <dinf> "]"
                        { "[P:" <paragraph number> “]"
                        { "[S:" <sentence number> "]"
                          <sentence>
                          "[/S]"
                         ["[RS]
                          <reference description>
                          "[/RS]"
                          "[DS]
                          <d_structure description>
                          "[/DS]"
                          ]
                        }...
                          "[/P]"
                         ["[RS]
                          <reference description>
                          "[/RS]"
                          "[DS]
                          <d_structure description>
                          "[/DS]"
                          ]
                        }...
                        "[/D]"
```

Descriptive format of <reference description> is as follows:

```
<reference description>   ::= { <referent node> "," <referee node> }
                              ...
<referent node>           ::= <uw node1> | <sentence node> |
                              <paragraph node>
<referee node>            ::= <uw node2> | <sentence node> |
                              <paragraph node>
<uw node1>                ::= { <UW>":"<UW-ID> | ":"<Scope-ID> }
                              [":"<sentence node>]
<uw node2>                ::= { <UW>":"<UW-ID> | ":"<Scope-ID> }
                              ":"<sentence node>
                              <sentence node> can not be omitted.
<sentence node>           ::= ":S:"<sentence number>
<paragraph node>          ::= ":P:"<paragraph number>
```

Descriptive format of <d_structure description> is as follows:

```
<d_structure description>   ::= { <relation> “(“ <sentence
                                  node>|<paragraph node> ","   <sentence
                                  node>|<paragraph node> ")" } ...
```

3. What is UNL

The UNL is a language for computers. As a language for representing information and knowledge described in natural languages, UNL has all the components corresponding to that of a natural language. It is composed of words expressing concepts called "Universal words", also referred to as UWs that are inter-linked with other UWs to form the UNL expressions of sentences. These links, called as "relations", specify roles of each word in a sentence. The subjective meanings intended by the author are expressed through "attributes".

3.1 Relations

There are many factors to be considered in choosing an inventory of relations between concepts. Different factors taken into account in choosing the relations lead to different sets of the relations. The UNL relations are selected according to the following principles.

PRINCIPLE 1 : NECESSARY CONDITION
When an UW has relations between more than one other UWs, each relation label should be set so as to be able to identify each relation on the premise that there is enough knowledge about the concept of each UW expressed.

PRINCIPLE 2 : SUFFICIENT CONDITION
When there are relations between UWs, each relation label should be set so as to be able to understand the role of each UW only by referring to the relation label.

The following are the relations defined according to the above principles. A relation label is represented as strings of 3 characters or less.

agt	Agent	a thing in focus that initiates an action
and	conjunction	a partner to have conjunctive relation
aoj	thing with attribute	a thing that is in a state or has an attribute
bas	Basis	a thing used as the basis (standard) of comparison
ben	beneficiary	an indirectly related beneficiary or victim of an event or state
cag	co-agent	a thing not in focus that initiates an implicit event that is done in parallel
cao	co-thing with	a thing not in focus that is in a parallel

	attribute	state
cnt	content	the content of a concept
cob	effected co-thing	a thing that is directly affected by an implicit event done in parallel or an implicit state in parallel
con	condition	a non-focused event or state that conditions a focused event or state
coo	co-occurrence	a co-occurrent event or state for a focused event or state
dur	duration	a period of time during which an event occurs or a state exists
equ	equivalent	an equivalent concept
fmt	Range	a range between two things
frm	Origin	an initial state of a thing or a thing initially associated with the focused thing
gol	goal/final state	a final state of object or a thing finally associated with the object of an event
icl	included/ a kind of	an upper concept or a more general concept
ins	instrument	a instrument to carry out an event
int	intersection	a partner to take an intersection
iof	an instance of	a class concept that an instance belongs to
man	manner	a way to carry out an event or characteristics of a state
met	method	means to carry out an event
mod	modification	a thing that restricts a focused thing
nam	Name	a name of a thing
obj	effected thing	a thing in focus that is directly effected by an event or state
opl	effected place	a place in focus affected by an event
or	disjunction	a partner to have disjunctive relation
per	proportion, rate of distribution	a basis or unit of proportion, rate of distribution
plc	Place	the place where an event occurs, or a state is true, or a thing exists
plf	initial place	the place where an event begins or a state becomes true
plt	final place	the place where an event ends or a state becomes false
pof	part-of	a concept of which a focused thing is a part
pos	possessor	a possessor of a thing
ptn	partner	an indispensable non-focused initiator of an action
pur	purpose or objective	the purpose or objective of an agent of an event or the purpose of a thing which exists
qua	quantity	Quantity of a thing or unit
rsn	reason	a reason why an event or a state happens
scn	Scene	a scene where an event occurs, or state is true, or a thing exists
seq	sequence	a prior event or state of a focused event or state
src	source/initial state	the initial state of an object or thing initially associated with the object of an event
tim	Time	the time an event occurs or a state is true
tmf	initial time	the time an event starts or a state becomes true
tmt	final time	the time an event ends or a state becomes false
to	destination	a final state of a thing or a final thing (destination) associated with the focused thing
via	intermediate place or state	an intermediate place or state of an event

3.2 Attributes

Attributes are mainly for the purpose to describe the subjectivity information of sentences. They show what is said from the speaker's point of view: how the speaker views what is said. This includes phenomena technically called "speech acts", "propositional attitudes", "truth values", etc. Attributes are also used to express the range of concepts such as the concept indicate generic type of concept and so forth. This time, we newly introduce attributes to express logical expressions in order to strengthen the expressibility of the UNL.

Relations and UWs are used to describe the objectivity information of sentences. Attributes modify UWs or semantic networks (scope: compound concepts) to indicate subjectivity information such as about how the speaker views these states-of-affairs and his attitudes toward them and to indicate the property of the concepts. Attributes are divided into the following eight groups:

1. Describing logicality of UW

In UNL2005, we introduced new attributes to express logical expressions in order to strengthen the expressibility of the UNL, although the role of those attribute can be expressed by using UWs as ordinal UNL expressions.

@transitive	attached to an UW that has transitivity
@symmetric	attached to an UW that has symmetricity
@identifiable	attached to an UW that can identify the subject
@disjointed	attached to an UW or a group of UWs to show that all element concept do not hold common instance. All connected UWs do not share instances.

2. Describing times with respect to the speaker

@past	happened in the past
@present	happening at present
@future	will happen in future

3. Describing speaker's view on aspects of event

@begin	beginning of an event or a state
@complete	finishing/completion of a (whole) event
@continue	continuation of an event
@custom	customary or repetitious action
@end	end/termination of an event or a state
@experience	experience
@progress	an event is in progress
@repeat	repetition of an event
@state	final state or the existence of the object on which an action has been taken

These attributes are used to modify the attributes above, to express a variety of aspects of natural languages.

@just	Expresses an event or a state that has just begun or ended/completed
@soon	Expresses an event or a state that is about to begin or end/completed
@yet	Expresses an event or a state that has not yet started or ended/completed, together with *@not*

4. Describing speaker's view of reference to concepts

@generic	generic concept
@def	already referred
@indef	Non-specific class
@not	complement set
@ordinal	ordinal number

5. Describing speaker's view of emphasis, focus and topic

@contrast	Contrasted UW
@emphasis	Emphasized UW
@entry	Entry or main UW of a sentence or a scope
@qfocus	Focused UW of a question
@theme	Instantiates an object from a different class
@title	Title
@topic	Topic

6. Describing speaker's attitudes

@affirmative	Affirmation
@confirmation	Confirmation
@exclamation	Exclamation
@humility	In a humility manner
@imperative	Imperative
@interrogative	Interrogation
@invitation	Inducement
@polite	Polite way
@request	Request
@respect	Respectful way
@vocative	Vocative

7. Describing speaker's feelings and judgments

@ability	Ability, capability of doing something
@get-benefit	Speaker's feeling of receiving benefits through the fact or result of something (to be) done by somebody else
@give-benefit	Speaker's feeling of giving benefits by doing something for somebody else
@conclusion	Logical conclusion due to a certain condition
@consequence	Logical consequence
@sufficient	Sufficient condition
@consent	Consent feeling of the speaker about something
@dissent	Dissent feeling of the speaker about something
@grant	To give/get consent/permission to do something
@grant-not	Not to give consent to do something
@although	Something follows against [contrary to] or beyond expectation
@discontented	Discontented feeling of the speaker about something
@expectation	Expectation of something
@wish	Wishful feeling, to wish something is true or has happened
@insistence	Strong determination to do something
@intention	Intention about something or to do something
@want	Desire to do something
@will	Determination to do something
@need	Necessity to do something
@obligation	Obligation to do something according to (quasi-) law, contract, or …
@obligation-not	Obligation not to do something, forbid to do something according to (quasi-) law, contract or …
@should	To do something as a matter of course
@unavoidable	Unavoidable feeling of the speaker about doing something
@certain	Certainty that something is true or happens
@inevitable	Logical inevitability that something is true or happens
@may	Practical possibility that something is true or happens
@possible	Logical possibility that something is true or happens
@probable	(Practical) probability that something is true or happens

@rare	Rare logical possibility that something is true or happens
@unreal	Unreality that something is true or happens
@admire	Admiring feeling of the speaker about something
@blame	Blameful feeling of the speaker about something
@contempt	Contemptuous feeling of the speaker about something
@regret	Regretful feeling of the speaker about something
@surprised	Surprised feeling of the speaker about something
@troublesome	Troublesome feeling of the speaker about the occurrence of something

8. For convention

@passive	passive form
@pl	more than one
@angle_bracket	< > are used
@brace	{ } are used
@double_parenthesis	(( )) are used
@double_quote	“ ” are used
@parenthesis	( ) are used
@single_quote	‘ ’ are used
@square_bracket	[ ] are used

3.3 Universal Words

Universal Words are words that constitute the vocabulary of UNL. A Universal Word (UW) is not only a unit of the UNL syntactically and semantically for expressing a concept, but also a basic element for constructing a UNL expression of a sentence or a compound concept. Such a Universal Word is represented as a node in the semantic network of UNL expression.

3.3.1. Concept of UW

How to define a concept depends on how the concept is to be used in a knowledge representation (or processing) system, because it may affect the efficiency or power of the system. In UNL, the function of a concept is focused on in defining the concept. The function of a concept means that what semantic relations the concept can have with other concepts, or with what concepts the concept can have the semantic relations, or what roles the concept can fill in satisfying various semantic relations. From this point of view, possible relations that a concept can have with other concepts are adopted in defining the concept. The possible relations imply also that the partner concepts of these relations must be satisfactory. These possible relations not only make a concept identifiable but also describe the behavior of the concept. This definition of a concept provides information on how the concept should be used in a UNL expression.

Based on this, concepts of UWs are designed and divided into categories of nominal concepts, verbal concept, adjective concepts and adverbial concepts.

Nominal concept is followed by 'thing' immediately. Under 'thing', seven categories are provided according to types of concepts. They are “abstract thing”, “concrete thing”, “functional thing”, “place(icl>thing)”, “pronominal thing”, “time(icl>thing)” and “volitional thing”. **Verbal concepts** are divided into three categories represented by 'be', 'do' and 'occur'. **Adjective concepts** are divided into two categories according to the features of predicativity and attributivity. Attributive concepts are divided into further two categories, one consists of the attributive concepts that express quantity, and the other one consists of the attributive concepts that qualify a concept. These three categories are represented by 'uw(aoj>thing)', 'uw(mod<thing)' and 'uw(qua<thing)'. **Adverbial concept** is followed by 'how' immediately. The adverbial concepts that take 'obj' relation is defined by 'how({icl>how,}obj>thing)'.

3.3.2. UW

A UW is made up of a character string (an English-language word) followed by a list of constraints.

```
<UW>      ::= <headword> [<constraint list>]
```

Headword of a UW is an English expression (a word, a compound word, a phrase or a sentence) that is interpreted as a label for a set of concepts: the set made up of all concepts that may correspond to that in English. **Constraint list** restrict the concept of a UW to a subset or to a specific concept included within the Basic UW.

A UW can express various levels’ concepts depending on the restrictions given by constraint list. The UWs are divided into four types:

Basic UWs are character strings that correspond to English words. Such a basic UW denotes all the concepts that may correspond to those in English. However a basic UW is not used if the English expression is ambiguous. Such a basic UW is usually used as the headwords of Restricted UWs for its various specific concepts. A basic UW is used if the English expression has no ambiguity.

Restricted UWs are by far the most important. A Restricted UW is made up of a headword (English expression) with restrictions. It is necessary when the English expression of headword has broader sense (more meanings) than the concept aimed to define. The restrictions restrict the range of the concept that an English expression represents. Each Restricted UW made from an English expression represents a more specific or particular concept, or a subset of the concepts of the English expression. For example, following are the Restricted UWs made from the English word "state":

state((icl>express(agt>thing,gol>person,obj>thing)) is a more specific concept that denotes an action in which humans express something.
state(icl>country) is a more specific sense of "state" that denotes a country.
state(icl>region) is a more specific sense of "state" that denotes a region of a country.
state(icl>abstract thing) is a more specific sense of "state" that denotes a kind of condition that persons or things are in.
state(icl>government) is a more specific sense of "state" that denotes a kind of government.

The information in parentheses is the constraint list and it describes some conceptual restrictions; this is why they are called Restricted UWs.

Extra UWs (or **Imported Uws**) denote concepts that are not found in English and therefore have to be introduced as extra categories. Foreign-language words are used as headwords using English (Alphabetical) characters. For example, following are the examples of Extra UWs:

ikebana(icl>flower arrangement) is "a kind of flower arrangement" for the meaning of "something you do with flowers",
samba(icl>dance) is "a kind of dance", and
souffle(icl>food) is "a kind of food".

Temporary UWs A is not necessary to define, such as a number or an address of email. They only appear in a UNL document.

4. UNL Knowledge Base

The UNLKB is a semantic network comprising every directed binary relation between UWs. All binary relations of the UNL KB are in the following format: 'relation(UW1, UW2)=c', where 'c' is the degree of certainty, which has the value 0 (impossible) or from 1 to 128 (certain). This binary relation means "UW1 takes UW2 as the relation in certainty value c"; or "UW2 plays the role specified by the relation to UW1 in certainty value c".

4.1. Roles of the UNLKB

- The UNLKB Defines Semantics of **UWs**

A UW is a label for a concept. Concepts labeled by UWs are defined by describing the set of possible relations that each concept can have with other concepts in UNLKB. Definitions of possible relations of a concept with other concepts describe the behavior of the concept. This behavior is the property of a concept in the sense that the descriptions of behavior characterize the concept and provide enough information for understanding the semantic structure of a sentence which include the concept.

- The UNLKB Provides **Linguistic Knowledge** of Concepts

The behavior of a concept is considered as linguistic knowledge on the concept. This knowledge is used to provide semantic structure of sentences of natural languages. For example, an "author" is a "person", who can take various actions that a person can take, such as writing something and something might be a book, and so forth. This level of knowledge is necessary to provide the semantic background of natural language sentences. Further knowledge, for example real world knowledge, will be established based on this linguistic knowledge, using the UWs.

In the UNLKB, the semantics of UWs are defined using the UW system and linguistic knowledge of concepts is provided also based on the UW System.

4.2. UW System

In the UNL KB, all UWs are linked with each other through 'icl' (subclass), 'iof' (element/instance), or 'equ' (equivalent) relations. 'Icl' links a UW of a subclass concept to the class concept UW; 'iof' links a UW expressing an instance to a UW of a class concept; and 'equ' links a UW to an equivalent UW. The UWs related to each other through 'icl', 'iof' and 'equ' relations make up a **hierarchy** of UWs. This hierarchy of UWs is the UW system. This UW system allows having **multiple** super-class concepts. Accordingly, the UW system is a **lattice** type of network.

4.3. Features of the UNLKB

The hierarchy of the UW system is constructed by taking the **property inheritance** and **replacement** by super-class concept mechanisms into consideration. In UW system, lower UWs inherit the properties of upper UWs; and upper UWs can replace lower UWs to convey a more general sense in a specific context of the lower UWs. All these inheritance and

replacement are carried out through the relations 'icl', 'iof' and 'equ'.

In the UNLKB, all possible relations, such as 'agt', 'obj', etc, that an UW can have with others are defined for each UW. Every possible relation is defined between the two most general UWs of the two categories (of lower UWs) that can have the relation. Utilizing the property inheritance mechanism of the UW system possible relations of lower concepts are deductively inferred, and this inference mechanism can reduce the number of binary relations.

Replacement of lower UWs by upper UWs can cause problems by introducing ambiguities if the upper UWs are not close enough in meaning to the lower UWs. To avoid this, the upper UWs must be the closest UWs among all of the more general UWs . In other word, every UW must be positioned under the closest upper UWs.

4.4. Uses of the UNLKB

The UNLKB defines the syntax and semantics of the UNL. Such UNLKB is used in sentence analysis for disambiguation and in sentence generation for finding more general concepts when encountering a unknown concept to a target language. The UNLKB also is used to verify UNL expressions since it provides syntax and semantics of the UNL.

To fully utilize the functions of the UNLKB, all UWs (concepts) must be defined in the UNLKB. For convenience, the following templates are provided for defining UWs that express instances. With these templates, a UW that has the same restriction as one of these templates is not necessary to be defined in the UNLKB, and the corresponding template is used instead in referring to the UNLKB. For example, 'uw(iof>person)' is the template for 'John(iof>person)'.

5. Knowledge Representation in UNL

There are various kinds of knowledge written in natural languages. Such kinds of knowledge must be provided in appropriate way for humans and also for computers. Especially for computers, knowledge must be represented in UNL in the form available for reasoning by computers.

5.1. Three Levels of Knowledge

We consider that knowledge can be divided into three levels: linguistic knowledge, definitions of concepts, and others.

(1) Linguistic knowledge is the knowledge providing semantic information of words for understanding sentences of natural languages. For example, an “author” is a “person”, which can take various actions that a person can take. This level of knowledge is necessary to provide semantic background of natural language sentences.

(2) Definitions of concepts provide the knowledge of the concepts in connection with other concepts that can specify the concepts. For example “a person who write a book” is the sentence to define what is an “author”. Through this definition, it is possible to know that a person who wrote a book will be the author of the book. This level of knowledge is indispensable for reasoning in information retrieval. etc..

(3) The third level of knowledge includes every kind of knowledge provided by natural language sentences such as encyclopedia. This

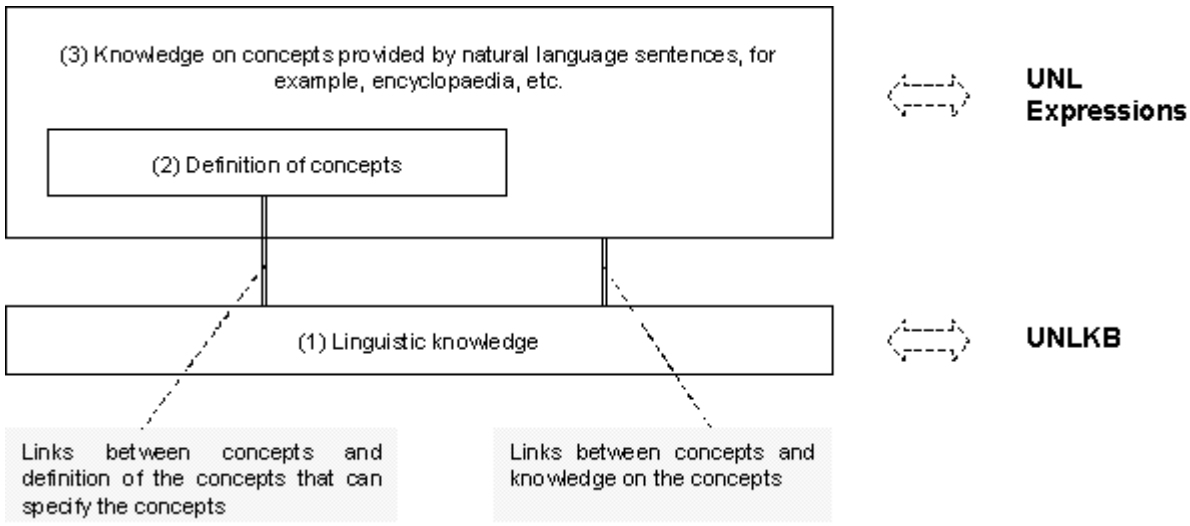

Figure 5. 1

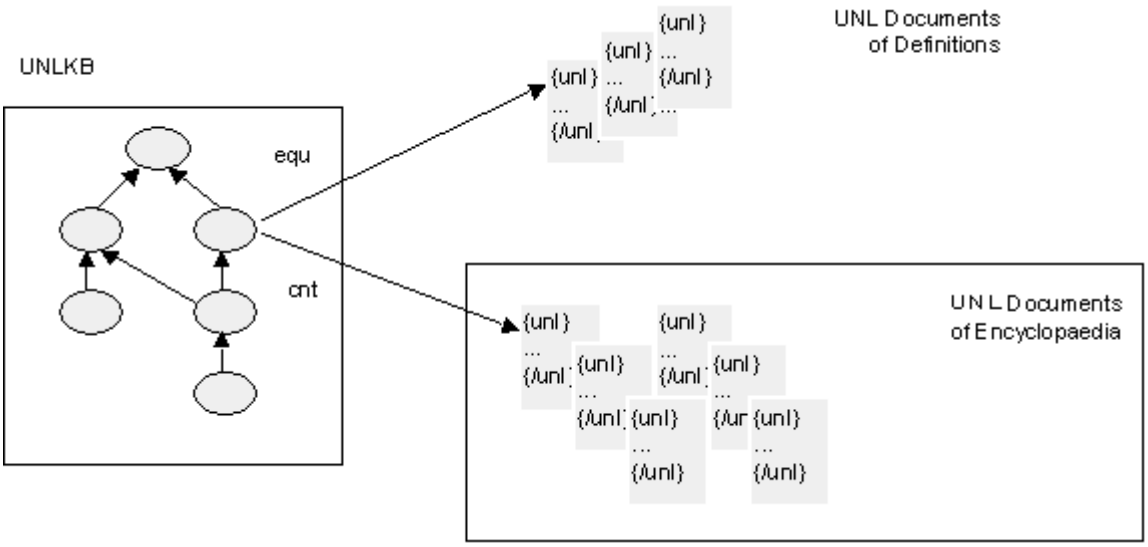

Figure 5. 2 Representation of the three levels of knowledge in the UNL Knowledge System

level of knowledge is for providing human with necessary information of concepts, and for providing computers with various information and knowledge on concepts necessary for reasoning and providing information on the concepts.

In the UNL system, linguistic knowledge is provided in UNLKB with UWs representing concepts. Definitions of concepts and knowledge on concepts described in natural language sentences are represented by UNL expressions. The following figure shows the relations of the three levels of knowledge and how they are represented in the UNL Knowledge System.

5.2. Representation of the Three Levels of Knowledge

The above figure shows how the three levels of knowledge are represented in UNL and how they are linked with each other in the UNL Knowledge System.

(1) UWs representing concepts of words are defined in UNLKB. The circles of nodes indicate UWs and arrows indicate relations. Concepts of UWs are defined in the UW System of the UNLKB. The UW System is a lattice type of hierarchy made of UWs related to each other.

(2) A definition of a concept (UW) is described in a file. Each file consists of sentences to define the concept of the UW. Sentences are expressed in UNL. Each definition file is linked with corresponding UW using relation 'equ'. All UWs used in the UNL expressions of definition must be defined in UNLKB.

(3) UNL encyclopedia is composed of files of sentences (UNL documents) describing information and knowledge about UWs. A file contains information and knowledge about a UW. Information and knowledge about a UW can be provided in different files. Each file of encyclopedia is differentiated using the developer name of the content as the file name. These files are linked with corresponding UWs using relation 'cnt'. All UWs included in UNL expressions of information and knowledge must be defined in UNLKB.

UWs are linked with definition files or encyclopedia files in the following format:

equ (<UW>, <*UNL Document of Definition*>)
cnt (<UW>, <*UNL Document of Encyclopedia*>)

Where,

<UW> ::= a UW defined in the UNLKB
<UNL Document of Definition> ::= a local file or a URL of a UNL document
<UNL Document of Encyclopedia> ::= a local file or a URL of a UNL document

A UNL document means a text file or a HTML file in which UNL expressions for the original natural sentences are included. A UNL document is attached to with extension ".unl" as in "http://www.undl.org/UNLencyclopedia/GreatWall.html.unl" or "x:/unlencyclopedia/local/GreatWall.html.unl".

6. Logic Expressions in UNL

The functions to express logic in UNL and the mechanism of reasoning using UNL are essential to establish Knowledge Infrastructure based on UNL.

6.1. Expression of Variables in UNL

Every concept is represented using UW in UNL expressions. Inference rules and necessary logical expressions are necessary to be expressed in UNL. Variables are necessary to write rules and logical expressions. UNL should deal with the variables. In UNL, variables are expressed as temporary UWs with variable names as headwords, the domains of variables as constraints of the UWs in the following format:

<variable name>"(iof>"<domain name>")"

Where,

<variable name>	::= a variable name such as "x", "y", etc
<domain name>	::= a UW that can express the domain. Full expression of a UW must be used.

For example, 'x(iof>dog(icl>canine))' is a variable expressing any element belonging to 'dog(icl>canine)'.

[**ex1**] Logical expression for "dogs are clever" is the following as shown in (1):

$\forall x$ (dog(x) -> clever(x)) (1)

UNL expression of this logical expression is the following as shown in [UNL1]:

```
[UNL1]
aoj:01 ( dog(icl>canine).@entry, x(iof>dog(icl>canine)) )
con ( :02.@entry, :01 )
aoj:02 ( clever(aoj>thing).@entry, x(iof>dog(icl>canine)) )
```

In case of logical expression "$\forall$x,y (x=parent(y) -> y=child(x))", more than one variables are necessary to introduce.

[**ex2**] For Logical expression:

$\forall$x,y (x=parent(y) -> y=child(x)) (2)

the UNL expression is shown as [UNL2]:

```
[UNL2]
aoj:01 ( parent(icl>relative).@entry, x(iof>parent(icl>relative)) )
pos:01 ( parent(icl>relative).@entry, y(iof>child(icl>relative)) )
con ( :02.@entry, :01 )
aoj:02 ( child(icl>relative).@entry, y(iof>child(icl>relative)) )
pos:02 (child(icl>relative).@entry, x(iof>parent(icl>relative)) )
```

In UNL logical expressions of [UNL1] and [UNL2], relation 'con' has the same function as the logical operator '->' (means "implication").

6.2. UNL Expressions for Logic and Meaning Representation

Logical expression in UNL [UNL1] describes literally "if x is a dog, it is clever". However, semantic expression of "dogs are clever" in UNL expresses the same meaning. This means that "if x is a dog, it is clever" is equivalent to "dogs are clever", and [UNL1] is equivalent to the [UNL1-1] below.

```
[UNL1-1]
aoj ( clever(aoj>thing).@entry, dog(icl>canine) )
qua ( dog(icl>canine), all(qua<thing) )
```

[UNL1] is more appropriate for reasoning, whereas [UNL1-1] is more simple and natural, and directly reflects the semantic structure of natural languages. Thus, [UNL1-1] is more appropriate for natural language analysis and generation. Generally, knowledge provided by natural languages is expressed by UNL in the simple way like [UNL1-1]. However, sometimes the logical expression, for example [UNL1], is necessary to be extracted from knowledge representation of UNL in order to facilitate reasoning process. Likewise, logical expressions resulted from reasoning is necessary to be converted into simple UNL expressions so that more natural sentences can be generated easily.

For instance, conversion between [UNL1] and [UNL1-1] can be carried out in the following steps:

Steps of conversion from [UNL1] to [UNL1-1]:

- 'x(iof>dog(icl>canine))' is an variable in the condition part 'aoj(dog(icl>canine), x(iof>dog(icl>canine)))' and the relation is 'aoj', we can get 'x(iof>dog(icl>canine)) = dog(icl>canine)'.
- In addition to this, 'x(iof>dog(icl>canine))' is a variable and can have all the element of 'dog(icl>canine)', 'qua(dog(icl>canine), all(qua<thing))' is added. Then substitutes the variable 'x(iof>dog(icl>canine))' for the value of the variable 'dog(icl>canine)', we can get 'aoj(clever(aoj>thing).@entry, dog(icl>canine))'.

Steps of conversion from [UNL1-1] to [UNL1]:

- Introduce a variable 'x(iof>dog(icl>canine))' to represent all the element of 'dog(icl>canine)' based on the binary relation 'qua(dog(icl>canine), all(qua<thing))' and make a condition part 'aoj:01(dog(icl>canine).@entry, x(iof>dog(icl>canine)))' and also make conclusion part 'aoj:02(cleaver(aoj>thing).@entry, x(iof>dog(icl>canine)))', then generate linking 'con(:02.@entry, :01)'.

6.3. UNL Expression on Quantification of Logical Expression

UWs basically represent class concepts or instances. UWs representing class concepts can be used to express all instances of the class, a subclass concept, or particular instances with quantifier. In UNL, quantifiers are expressed by UWs or attributes. The following tables show the attributes and UWs that have the function of indicating quantifiers which can restrict a UW representing a class concept to express a specific set of the concepts.

Attributes that have the functions for quantification:

@generic	Attached to a UW to express all instances of the concept
@def	Attached to a UW to express some particular instance of the concept
@indef	Attached to a UW to express existence of some instance of the concept. Corresponding to "existential quantifier" of logical expression
@not	Attached to a UW to express the complement concepts or instances
@ordinal	Attached to a UW to express some particular instance of the concept

Examples of UWs expressing quantities:

all(qua<thing)	Has 'qua' relation with a UW to express all instances of the concept. Corresponding to "universal quantifier" of logical expression
every(qua<thing)	Has 'qua' relation with a UW to express all instances of the concept. Corresponding to "universal quantifier" of logical expression
some(mod<thing)	Has 'mod' relation with a UW to express existence of some instance of the concept. Corresponding to "existential quantifier" of logical expression
some(qua<thing)	Has 'qua' relation with a UW to express quantity of instance(s) of the concept
no(qua<thing)	Has 'qua' relation with a UW to express no instance of the concept exists
most(qua<thing)	Has 'qua' relation with a UW to express most of instances of the concept
many(qua<thing)	Has 'qua' relation with a UW to express existence of many instances of the concept
<a number>	Has 'qua' relation with a UW to express the number of instances of the concept

The followings are examples to show how the universal quantifier($\forall$) and existential quantifier($\exists$) of logical expressions are expressed in UNL.

[**ex3**] Logical expression for "all is walking" is shown as in (3):

$\forall x$ (person(x) -> walk(x)) (3)

[UNL3] shows the logical expression, and [UNL3-1] shows the UNL semantic expression of the sentence.

```
[UNL3]
aoj:01 (person(icl>human).@entry, x(iof>person(icl>human)))
con ( :02.@entry, :01 )
agt:01 ( walk(agt>thing).@entry, x(iof>person(icl>human)) )
[UNL3-1]
agt ( walk(agt>thing).@entry, person(icl>human) )
qua ( person(icl>human), all(qua<thing) )
```

[**ex4**] Logical expression for "someone is walking" is shown as in (4):

$\exists x$ (person(x) $\wedge$ walk(x)) (4)

[UNL4] shows the UNL logical expression, and [4-1] shows the UNL semantic expression of the sentence.

```
[UNL4]
aoj:01 ( person(icl>human).@entry, x(iof>person(icl>human)))
int ( :02.@entry, :01 )
agt:01 ( walk(agt>thing).@entry, x(iof>person(icl>human)) )
[UNL4-1]
agt ( walk(agt>thing).@entry, person(icl>human).@indef )
```

Relation 'int' of UNL has the same function as intersection ($\wedge$) of logical operator.

Scope (domain) of Quantification

"everyone loves someone" has two interpretations: 1) every person has some person to love, and 2) all persons love the same one person. (5) and (6) show the logical expressions respectively:

$\forall x \exists y$ love (x, y) (5)
$\exists y \forall x$ love (x, y) (6)

As shown in (5) and (6), scopes of "universal quantifier" and "existential quantifier" are different

for the two meanings. In (5), "existential quantifier" exists within the scope of "universal quantifier", whereas in (6), "universal quantifier" exists within the scope of "existential quantifier".

In UNL, this scope of quantification can be expressed by either scopes (compound concepts) of UNL expression, or UWs or attributes that can quantify the different meanings. [UNL5] and [UNL6] show the UNL expressions, figures 6.1 and 6.2 show the graphs of the two different meanings of 1) and 2) using scopes.

In UNL, except cooperative concepts, **different initiators** are considered to **initiate different events**. For example "two dog are running" implies two independent events such as "one dog is running" and "another dog is running". According to this **interpretation**, the first meaning of 1) can be expressed through [UNL5] and the second meaning of 2) can be expressed through [UNL6].

In [UNL5], "love someone" is expressed in a scope. Within the scope, "someone" is linked to "love" by 'obj'. "every person" as agent is linked to the scope. According to the interpretation of different agent initiating different event, the number of events exist according to the number of agents, the initiators of the events respectively. In each event, the "person" to love exists for each event through relation 'obj'.

In [UNL6], "everyone loves" is expressed in a scope. Within the scope, "everyone" is linked to "love" as agent. "someone" is linked to the scope as object. According to interpretation mentioned above, different events caused by different agents (initiators) exist inside the scope. All these events are linked by one relation 'obj' to indicate the same "person" to love.

```
[UNL5]
agt(:01.@entry,    person(icl>human):01)
qua(person(icl>human):01,    every(qua<thing))
obj:01(love(agt>thing,obj>thing).@entry,    person(icl>human):02)
mod:01(person(icl>human):02,    some(mod<thing))
[UNL6]
agt:01(love(agt>thing,obj>thing).@entry,    person(icl>human):01)
qua:01(person(icl>human):01,    every(qua<thing))
obj(:01.@entry,    person(icl>human):02)
mod(person(icl>human):02,    some(mod<thing))
```

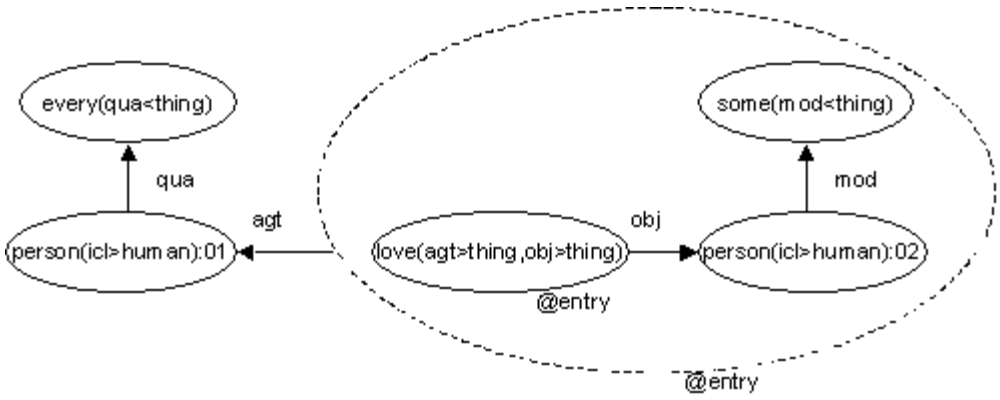

Figure 6. 1 Graph of [UNL5]

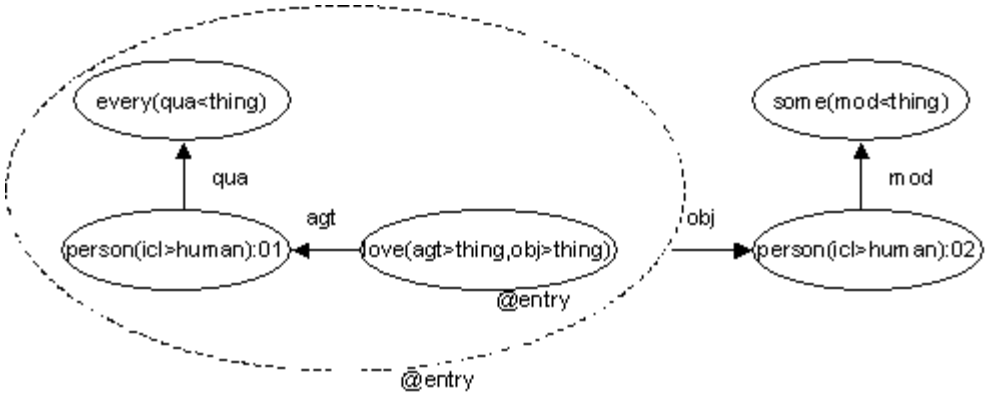

Figure 6. 2 Graph of [UNL6]

6.4 Attributes and Relations that can Express Logic

Attributes of UNL have been developed mainly to express subjective information with respect to the author of a sentence. In UNL2005, new attributes to express logical expressions are introduced in order to strengthen the expressiveness of the UNL, although the roles of those attributes can be expressed using UWs in ordinal UNL expressions. In this section, attributes and relations for expressing logic are summarized.

6.4.1. Attributes for Expressing Logic of UW

The following table shows the attributes for expressing logic of UW. These attributes are used to attach to the UWs that have the logic. ':@disjointed' can be attached to a group of UWs (a scope) to show that every concept included in the group are disjointed. It means that all concepts do not share instances.

@transitive	attached to an UW that has transitivity
@symmetric	attached to an UW that has symmetricality
@identifiable	attached to an UW that can identify the subject
@disjointed	attached to an UW or a group of UWs to show that all concepts do not hold common instances. All connected UWs do not share instances.

6.4.2. Relations for Expressing Logical Relations

These relations of UNL have been developed to focus on expressing semantics of a sentence. From this point of view, the relations can be divided into three groups: 1) to express case relations ('agt', 'obj', etc), 2) to express event relations ('con', 'seq', etc), and 3) to express conceptual relations ('cnt', 'equ', etc). The following table shows the relations that can be used to express logic and the meaning when they are used to express logic.

Relations that can be used to express logical relations between concepts	
And	Connects two concepts that values are true. Corresponds to the logical operator "union"
Con	Shows the conditional concepts. Corresponds to the logical operator "imply" but the direction is reverse
Int	Connects two concepts and takes an intersection. Corresponds to the logical operator "intersection"
Or	Connects two events that one of them is true at least. Corresponds to the logical operator "exclusive union"
Relations that can express logical relations	
cnt	Link a concept to the content
equ	Link to an equivalent concept
icl	Link to an upper class concept
iof	Link an instance to a class concept

7. UNL System

The UNL System consists of three major components: language resources, software for processing the language resources, and tools and systems for maintaining and operating the language processing software or language resources. Language resources are divided into language dependent part and language independent part. Linguistic knowledge on concepts that universal to every language is

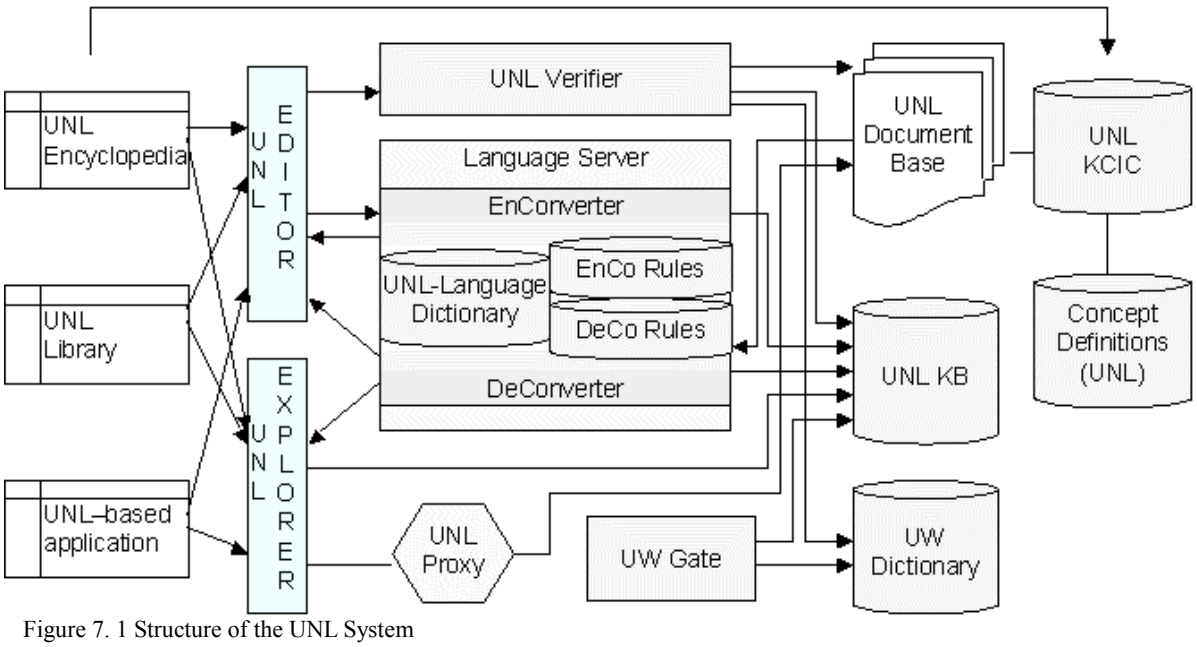

Figure 7. 1 Structure of the UNL System

considered language independent and to be stored in the common database UNLKB. Language dependent resources like word dictionaries and analysis and generation rules, as well as the software for language processing, are stored in each language server. Language servers are connected through the Internet. Supporting tools for producing UNL documents can be used in a local PC. Such supporting tools operate with consulting language servers through the Internet. Verification of UNL documents can be carried out through the Internet or in a local PC. UW Gate for searching and maintaining the common database UNLKB operates through the Internet.

7.1 Structure of the UNL System

In figure 7.1, highlighted parts show the components of the UNL System, white parts show applications and their UNL database based on the UNL System.

UNLKCIC stores information of Key Concept in Context (KCIC) about UNL documents. The UNLKCIC is made for every binary relation of UNL documents. This information is used for searching related UNL expressions of a UNL expression. Through UNLKCC, every UW of UNLKB is linked to the UNL documents each UW involved, and all UWs included in the UNL documents are also stored in the UNLKB and linked to corresponding UNL documents. UNLKCIC expands the knowledge in UNLKB from merely linguistic knowledge such as binary relations between context independent concepts to real world knowledge such as binary relations between context sensitive concepts restricted by other concepts knowledge.

Concept Definitions are the collection of UNL documents made for the sentences of definitions of UWs. These definitions of concepts provide the knowledge of concepts in connection with other concepts that can specify the concepts. This knowledge is indispensable for reasoning in information retrieval, etc.

UNL Documents mean the documents in which UNL expression is described for each sentence of natural language. A UNL document can be made of a plain text file or an UNL-embedded html file. A UNL document base is a collection of UNL document files. UNL documents are for the purpose to provide information and knowledge in UNL expression.

UNLKB is a semantic network comprising every directed binary relation between UWs. These binary relations are the possible relations that each UW can have with others. Such relations are established based on the UW System. Such UNLKB has the following functions: 1) defines semantics (concepts) of UWs, and 2) provides linguistic knowledge of concepts.

UW Dictionary stores and provides the interface between UWs and words of natural languages.

UNL Verifier verifies whether a UNL expression is correct syntactically, lexically and semantically. The syntax check of a UNL expression is carried out against the UNL Specifications. In lexical check, whether all UWs of a UNL expression are defined in the UNLKB are checked. In semantic check, whether each binary relation of a UNL expression is defined as possible is certified with consulting the UNLKB.

UNL Language Servers (LSs) are located in the Internet to carry out the conversions between natural languages and UNL expressions. Each LS contains an EnConverter and a DeConverter of a language. EnConverter converts natural language sentences to UNL expressions. DeConverter convert UNL expressions to natural language sentences.

UW Gate is a tool for people to access the UNLKB and the UW dictionary through the Internet.

UNL Proxy Server works in a local computer to communicate with language servers. It functions as a filter to check whether a web page that a user required is written in UNL or not. If UNL expressions are included in the web page, it communicates with an appropriate language server in the Internet for deconverting the UNL expressions into desired language sentences and provides the Internet browser with the results to display.

UNL Editor is a tool helping to produce UNL documents. It includes an EnConverter and a DeConverter. Each of them can be selected according to language. EnConverter converts natural language sentence into UNL expressions. DeConverter provides generated results as feedback for checking the correctness of UNL expressions.

UNL Explorer provides the basic means to knowledge infrastructure. It manages UNLKB, UNLKCIC and UNL documents and provides knowledge or information through UWs. The UNL Explorer can be used in two ways. For human, it allows users or developers to view or to develop the UNL Knowledge System such as the UNL Encyclopedia. For computers, it provides information or knowledge on UWs.

The UNL Explorer uses UNLKB for navigating information stored in UNL database. It has two windows: the hierarchy of UWs (UW System) of UNLKB is shown in the left window. UWs of the UNLKB are keys for information stored in UNL database. Information on UWs is shown in the right window through navigation through the UW System. Information on UWs is described in UNL documents.

All UWs used in the UNL documents are included in the left window of UNLKB and are keys for further information.

The UNL Explorer allows users to search for information using UWs or words of natural languages. It shows the information in UNL or a desired natural language by accessing UNL Language Servers. It also provides functions for developers to add information to or modify information of the UNL database in their native languages.

Information about a UW is stored in a file. Location of the file is linked with the UW. This architecture of the UNL Knowledge System allows its development to be carried out by a wide range of developers from different languages and cultures. Such a database can provide a wealth of up-to-date information on various aspects of information and knowledge from all over the world.

UNL Encyclopedia is a collection of UNL documents describing various knowledge or information. The merit of the UNL Encyclopedia is that it enables each people to use its native language to share knowledge or information with other peoples in their respective languages, through UNL Explorer for example. The UNL Encyclopedia basically describes ordinary encyclopedias in UNL and integrates those descriptions (UNL documents) into the UNLKB by linking the UNL documents to corresponding UWs. It is a multilingual encyclopedia navigating through the UNLKB for human, with the knowledge description necessary for computers to process knowledge.

UNL Library is an UNL-based application providing bibliographic information (e.g. MARC21) and contents (in future) in UNL, and allowing users to search for books through UNL.

7.2. Mechanism of conversion of UNL expressions

Figure 7.2 shows the mechanism how a UNL document is made and how a UNL document is converted into natural languages in the UNL system. Arrows in solid line show dataflow, arrows in broken line show access.

The EnConverter and DeConverter are the core software in the UNL system. The EnConverter converts natural language sentences into UNL expressions. The Universal Parser (UP) is a specialized version of the EnConverter. It generates UNL expressions from annotated sentences using the UW dictionary without using grammatical features. All UNL expressions are verified by the UNL verifier. The DeConverter converts UNL expressions to natural language sentences.

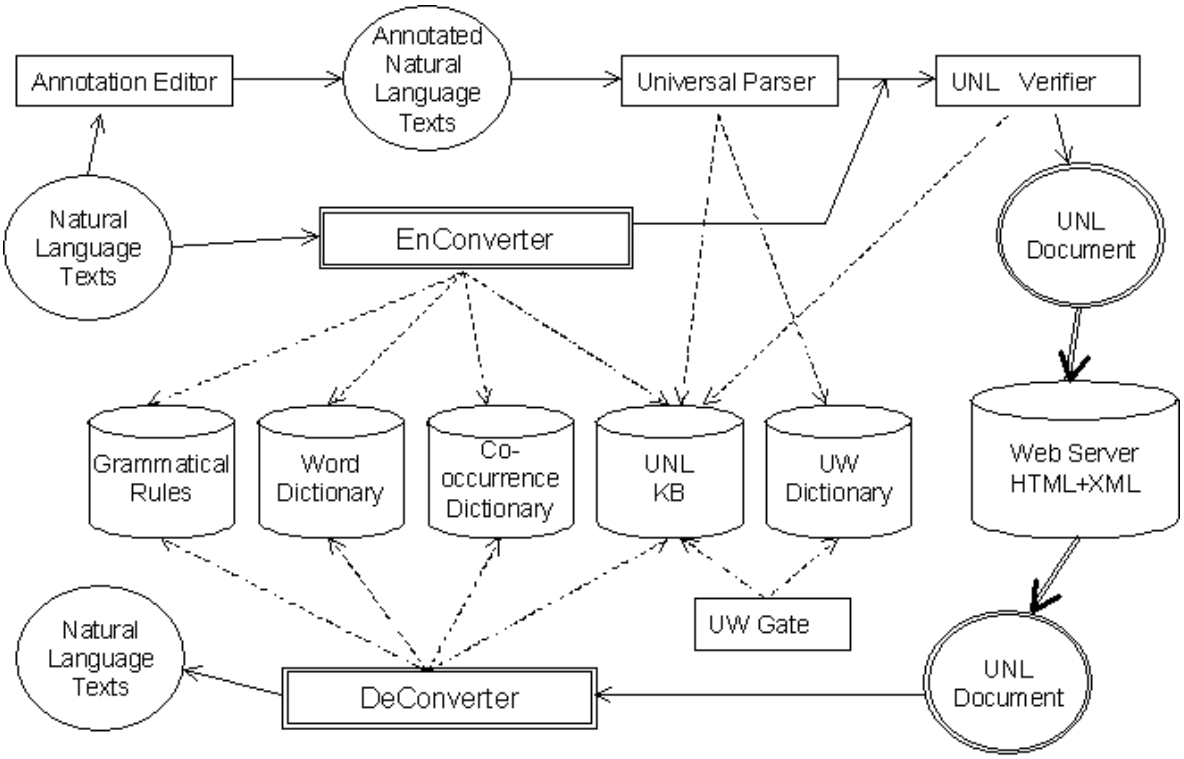

Figure 7.2 Mechanism of conversion of UNL

8. Conclusion

The UNL is a language for computer to express information and knowledge that can be described in natural languages. Methodology of defining UWs allows to express any specific concept of a particular language precisely without ambiguities and understandable for computers and for humans of other languages. This mechanism allow all languages to have interfaces with UNL and makes it possible that anyone can participate in the development of a language module or UNL-based applications. Information and knowledge that UNL will treat in the future needs a wide range of people to develop them. The knowledge expressed by UNL is universal and can be shared by all nations.

With UNL, knowledge can be expressed understandable to computers. This means that computers can reason using the knowledge. The UNL is a language for providing the Knowledge Infrastructure for more intelligent processing such as semantic computing. The UNL expresses knowledge provided only by natural languages at this moment. Whereas CDL (Conceptual Description Language: an integration of UNL (Universal Networking Language), RDF, OWL, and so forth) is aimed at dealing with contents provided by all kinds of media.

Related materials and References

[1] UNDL Foundation: UNL System http://www.undl.org/unlsys/

[2] UNDL Foundation: UNL Specifications 2005 **http://www.undl.org/unlsys/unl/UNL2005/**, 2005

[3] UNDL Foundation: UNL Specifications v3.3 http://www.undl.org/unlsys/unl/UNLSpecs33.pdf, 2004

[4] UNDL Foundation: UNL explorer and UNL Encyclopedia http://www.undl.org/encyclopedia/, 2004

[5] Kinsui Satoshi and Ikumi Imani: Semantics and Context, Iwanami shoten, Introduction to contemporary linguistics 4, 2003