[MT Summit IV, July 20-22, 1993, Kobe, Japan]

Interlingua for Multilingual Machine Translation

UCHIDA Hiroshi*
FUJITSU LABORATORIES LTD.

ZHU Meiying
CENTER OF THE INTERNATIONAL COOPERATION
FOR COMPUTERIZATION

July 20. 1993

1 Introduction

In machine translation systems, an intermediate representation is necessary to express the result of sentence analysis. This represents syntactic structure or semantic structure of an input sentence given as a character string. The syntactic structure is normally represented is a tree and semantic structure as a network.

In transfer system, the intermediate representation of syntactic structure is usually the same as analyzed structure of input sentences. When the intermediate representation reflects the syntactic structure of a source language, syntactic structure transfer from a source language to a target language is necessary to translate. Implementation cost for the transformation becomes enormous if we develop a multilingual translation system with such an approach.

The intermediate representation of interlingua approach is called interlingua. It is the semantic expression of a sentence, how concepts expressed by each word of a sentence relate each other, and what role they play.

When we develop the multilingual machine translation systems of interlingua approach, this approach has two important merits.

The first merit is that the development of a machine translation system for a language can be localized. To develop dictionaries and grammar rules for analysis and generation of a language requires well-trained native speaker only for an interlingua. An analysis system and a generation system of an interlingua system can completely separated each other by the interface of interlingua. Therefore the development of an analysis system and generation system can be developed independently from those of the other language. Accordingly, those who develop an analysis system and a generation system only need to know the language and the interlingua.

The second merit is that knowledge necessary for machine translation systems can be utilized in common. To understand the sentence written in the natural language

This work was done when the author belonged to CICC.

both humans and computers must know the meaning of words and its usage. This kind of knowledge is very large-scale, however, without it, the natural language can not be analyzed as humans do. To do a high-quality machine translation semantic analysis using this knowledge is necessary. If this knowledge is described using interlingua, we can utilize it commonly in analysis systems of each language. This greatly gives merits to eliminate the waste that we develop the same knowledge in the different format.

Up to present, several kinds of interlinguas have been used in machine translation systems of interlingua approach. The main ones of those interlinguas are, CETA project studied mainly by Grenoble University in France, ATLAS (Fujitsu), PIVOT (NEC) and KBMT (Carnegie Mellon University). These interlinguas use the same framework in the following point: they express the meaning of a sentence using a symbol which represents a meaning or concept, and the relation between symbols by a certain means.

2 CICC's Interlingua

An interlingua must represents all the information which a, sentence expresses. A sentence generally has various kind of meanings. These meanings are represented in the form of a word itself, the combination, tense, aspect, mood and sentence style, but how to express these things, of course, varies depending on the language. An interlingua must represent all these kind of information in universal way.

CICC's interlingua vocabulary mainly consists of concepts which express the meaning of words and relations between concepts which express the meaning relations of words in a sentence. Concepts are expressed by headconcepts and relations are expressed by relation labels. In addition to them, we introduced attributes and special concepts.

The interlingua, is a, set of binary relations between concepts and unary relations. Binary relations express the relation between two headconcepts and the concept expressed by them. Unary relations express the concept expressed by a headconcept and an attribute. These concepts implied by the interlingua. restrict each other.

The binary relations between concepts indicate that "concept 1" and "concept 2" have a relation that is expressed by the "relation". The unary relation indicates the concept expressed by "concept" restricted by the "attribute". The interlingua is expressed by semantic networks that take headconcepts as nodes and relations as arcs.

Characteristics of this interlingua include the following: (1) It is made up of hyper networks which can take compound concepts as one concept or as concepts which are made up of more than two elemental concepts. (2) One concept is given to one entity so that it can be referred to from various levels.

A sentence expresses a (compound) concept. The concept of a sentence is expressed by a set of binary relations between concepts of constituent words express. The expression is the same as the expression of concept dictionary. This is because we aim at sharing knowledge which is necessary for machine translation. If the descriptive format of knowledge is the same as the framework of the interlingua. the analyzing system of each language can share the knowledge in common so that we can omit the redundancy of constructing the same kind of knowledge in another form for each system.

Tense, aspect, mood and sentence style information is the most language-dependent part, but if these informations are represented in the unchanged form, it may not be suitable to be called an interlingua. Because of this, these informations need to be represented arranging from the different standpoint which is common to every language. We introduced the frame to express speaker's perspective, speaker's intention, speaker's feelings or judgement and structure of a sentence.

Speaker's perspective for the concept of a sentence represents how and from where the speaker observe the concept. This information is expressed as a past, present, future tense or an aspect in a sentence. Speaker's intention to express the concept and speaker's feelings or judgement for the concept represent how they intend to issue or judge the concept of a sentence. This information is expressed as mood such as imperative, interrogation, supposition. Structure of a sentence is to reflect original sentence structure in the interlingua, which represents mainly relations between sub-sentences and headings or relations between previous and next sentence.

Correspondence of information of natural language sentence and interlingua are as follows.

natural language	interlingua
independent word	concept
connection of words	relation between concept
sentence style	attribute
tense and aspect	attribute

2.1 Concept

A concepts in the interlingua are expressed by a headconcept which is explanatory sentences of the concept. A concept indicated by a headconcept is a set of elements which have property set. Each property is indispensable to form the concept. The concept of a word is a set of common properties which does not depend on specific contexts or situations in which the word is used.

Headconcepts serve to distinguish the concepts expressed by words. Concepts expressed by words include general concepts (E.g.:"Birds", "Fly", etc.), those that express special concepts such as pronouns (E.g.:"I", "We", etc.), those that express direction or relative position such as prepositions and directors (For Chinese). (E.g.:"Before", "Top"). Apart from these, there are also special headconcepts which are expressed by postpositions and auxiliary verbs.

A headconcept of interlingua represents an instance of the concept represented by the headconcept. This headconcept represents unspecified elements of the concept unless any particular instruction. All of the elements are represented by attaching the attribute "all", and by attaching the attribute "generic", an intension of concept, is represented.

2.2 Relation

In the interlingua, relations between concepts are expressed by relation labels. These relations are for represent meaning relations of words to express compound concepts expressed by sentences.

We considered following in designing relations between concepts. The relations should be designed to be able to represent all compound concepts sentences express. And the structure of relations should be simple for easy treating of the interlingua.

According to this design principle, we did not introduce redundant relations by clarifying roles of relations, and also did not introduce the hierarchical system on relations for simplicity of the structure of the interlingua.

Relation set of the interlingua thus designed are case relations such as "agent", "object", "implement", pseudo relations such as "possessor", "from-to", and semantic relations such as "part of", "kind of", etc.

(1) General Relation Labels

agent	subject winch causes an action and event with intention
object	object influenced by an action
	subject of change
	object of existence
a-object	object having attribute (attribute object)
manner	way how an action is done, way how a state is changed
	way how state or relation is
cause	cause of an event
implement	a tool or method of action
material	material of an action or element of a state
time	time when an event occurs
time-from	time when an event begins
time-to	time when an event ends
source	original position of a subject or object in an event (action or change)
goal	last position of a subject or object in an event (action or change)
place	place (physical space) where an event occurs
scene	scene (logical space) where an event occurs
condition	condition of event occurs
cooccurrence	simultaneous (progress) relation of events
sequence	sequential order of events
quantity	quantity of a thing or amount of change in an event
number	number of a thing

basis	criteria of comparison
and	conjunctive relation of concepts
or	disjunctive relation between concepts
purpose	purpose of an act ion
modifier	restricts a concept

(2) Pseudo-relation Labels

Pseudo-relation labels have been introduced for simpler representation of concept relations. The pseudo-relation labels in representations can be replaced by normal relation labels by supplementing concepts.

possessor	owner of an object
from-to	range of a thing or an event
	unit of a thing or an event
beneficiary	beneficiary of an event (a person who receives a benefit or disadvantage)

(3) Semantic Relation Labels

part of	part and whole relation
kind of	successive relation

2.3 Attribute

We introduced attributes in interlingua vocabulary in addition to concepts and relations. Attributes are used to restrict concepts and to represent the perspectives and intention of speakers.

Since the concept described by a headconcept represents unspecified elements of the concept, some method is required to restrict them. Attributes are introduced to restrict the concept are such as "not" for negation, "begin" for beginning of event, "end" for ending of event. These attributes are used to restrict the concept of things or events.

(1) Attributes which restrict a range of concept

all	all elements that belong to a class defined by a concept
generic	intension of a concept
not	all concepts except a concept denied (complement of a concept

(2) Attributes which restrict aspects of an event

begin	beginning of an event
end	finish of an event
progress	event is going on
continue	repeating action or action continue
state	a state or result which are reached after finish of an action
complete	completion of an event

Attributes for representing the speaker's perspectives express information which is expressed by tense or aspect. Information expressed by tense is considered to indicate what point of time the speaker observes, on the other hand, information expressed as an aspect is considered to indicate how the speaker observes an event or fact from that point. Former attributes are such as "past", "present" and "future", and latter attributes are such as "yet"(not yet an event occurs), "already", "soon" and "just".

(3) Attribute for temporal location of perspective of the speaker

past	perspective of past
present	perspective of present
future	perspective of future

(4) Attributes which represent aspect information

yet	indicates that an event has not begun yet
already	indicates that an event has already begun
soon	indicates that an event are about to begin
just	indicates that an event has just begun

Attributes for representing speaker's intention express speaker's mental state or attitude when he utters a sentence. These informations come from mood or sentence form of sentences. These attributes are such as "imperative", "grant", "invite" and so forth.

(5) Attributes which represent speakers' intention for sentences

obligation	
imperative	make strong obligation based on subjective demand
request	make medium obligation based on subjective demand
invite	invitation
duty	make strong obligation based on objective demand
should	make medium obligation based on objective demand

advice	make weak obligation based on objective demand
threat	
natural-thing	ideal condition, natural course, common sense, custom
grant	
grant	grant
sufficiency	sufficiency
grant-not	must not
agreement	
natural-result	natural consequence of something
require agreement	
require-agreement	require agreement act
judgement	
if	supposition
reality	a reality
thought	subjunctive
probability	
maybe	inference of possibility
seem	inference, conjecture
sure	inference from situation (certainty)
rumor	rumor
appearance	condition, circumstance
belief	
conclude	assertion
feeling	
blame	blame
exclamation	admiration
pity	pity
regret	regret
underestimate	underestimation
unexpected	unexpectedness, unforeseen
feelings for benefit	
get-benefit	receives a benefit from agent
give-benefit	an agent gives a benefit to someone
interrogation	
interrogation	question
respect	
polite	politeness
respect	respect

(6) Attributes which represent speakers' intention for sentence elements

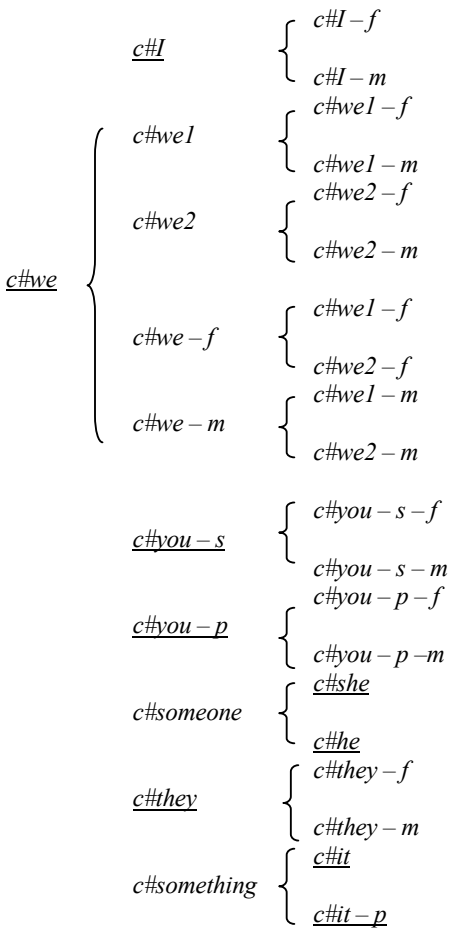
emphasis	emphasis
focus	focus
main	chief concept
topic	topic
wh	indicates the unspecified thing which speaker wants to know
come	to come close to the basis thought by the speaker
go	to go further from the basis thought by the speaker
specific	specific demonstrative
this	immediate demonstrative
that	distant demonstrative

2.4 Special Conceptual Representation

Concepts expressed by pronouns and concepts expresses relative concepts are different from general concepts. Concepts expressed by pronouns represent entities referred by pronouns such as speakers, listeners, or certain persons in certain situations. Relative concepts are linked closely to their objects. It is necessary to introduce the frame easy to treat these concepts.

We introduced special concepts (headconcepts) for easiness of reference and also we introduced basic set of these concepts. And we defined the relation (upper or lower) for these concept to clarify the position in the concept hierarchy.

According to the language, the way of expression of these concepts is quite different. If headconcepts are established directly for such concepts according to languages, many different kinds of headconcepts may be established for the same concept. Then the inter-ingua will become difficult to understand. For this reason, for fundamental concepts of these kinds, standard headconcepts are necessary. On the other hand, some of concepts which are expressed by pronouns contain information on if it is feminine or masculine (depending on languages), if it includes the listener or not, etc. Common fundamental concept sets must be given to such concepts. And then, for pronoun expressions peculiar to a language, individual headconcepts are established for the language and are positioned in the concept hierarchy to clarify their positions in respect to the fundamental concept set. Therefore to generate other languages for such individual concepts, either upper or lower concepts will be used. Concepts expressed by pronouns are expressed by special headconcepts. The following are the various types of pronoun concepts. Of these, those underlined are fundamental concept sets.



Remarks:

-f indicates female

-m indicates male

-p indicates plural

-s indicates singular

c#we1 includes “listener”

c#we2 does not include “listener”

For concepts which express direction or relative position, the fundamental concept set is established, and by combining them, compound concepts are expressed.

(1) Relative concepts expressing space

c#front	front
c# behind	behind
c#left	left
c# right	right
c#upper	above
c#lower	below

(2) Relative concepts expressing parts

c#front-part	front part
c#behind-part	back part
c#left-part	left part
c#right-part	right part
c#upper-part	upper part
c#lower-part	lower part

(3) Relative concepts expressing time

c#before	time before
c#after	time after

To express sentence structure and information on speaker's intentions and judgments for the overall sentence, we introduced a special headconcept "c#statement".

Sentences are generally made up of several sub-sentences such as complex or compound sentences. To express such sentences in the interlingua, a special concept "c#statement" is attached to the whole sentence as well as the various sub-sentences that make up the sentence to discriminate them. Information on relations between these sub-sentences are expressed by connecting the related sentence and its c#statement using relation labels which express sentence structure. There are six types of relations.

previous-st	previous sentence
sub-st	subordinate sentence
co-st	conjunctive sentence
quotation	quotational sentence
modify-st	modifying sentence
statement	to connect c#statement in the predicate concept, of a sentence to represent in the form which indicates the predicate concept

Information on the speaker's intentions and judgments is expressed by attaching the corresponding attributes to the special concept c#statement of a sentence. Several of these attributes can be attached to one c#statement concept.

We also introduced the null concept in the interlingua vocabulary. When there a re some omissions in a sentence, Making the interlingua for such a sentence, we need to supplement the concept which is correspond to the omitted concept. But when we restrict the role of the omitted concepts from the surrounding concepts and relations, we can use null concept instead of supplementing the corresponding concepts. In the tupple relation of c#X – f — c$f(X), if the concept c#f(X) to be supplemented is restricted by the concept c#X and relation f, it does not have to be supplemented in the interlingua. expressions containing t h e information c#X – f —. Null concept is indispensable to express various kind of sentences in the frame of the interlingua.

3 The Interlingua in Machine Translation

In machine translation, one of the problems is how to find correctly corresponding words. Both in transfer approach and interlingua approach, it is necessary to have bilingual dictionary of some kind. In transfer approach, a bilingual dictionary defines the correspondence between source language and target language words. In interlingua approach, a bilingual dictionary(often called monolingual dictionary) defines the correspondence between source language or target language and interlingua.

We designed the interlingua as a method of meaning representation which is not influenced by various languages phenomena.

Concepts obtained from source language must be represented in target language. When concepts are common to target language, corresponding words of target language are directly
retrieved from the dictionary of target language. But when concepts are not common to target language, using the concept hierarchy, the concept is converted to similar concepts of target language and then translated sentence is generated.

To realize this, in the concept hierarchy, all the concepts must be positioned .

The merit of t h i s system is that we develop a dictionary of a language independently not considering the corresponding words of other languages. And when we add a new language, we do not necessary to consider other languages, we only focus on the development of new language. Therefore, the development cost of adding new language will be low.

4 Maintenance of the Interlingua

To establish vocabulary of the interlingua. following methods are considered. First method is that, to establish interlingua vocabulary (neutral lexicon item) at the pivot of corresponding words of all target languages from multilingual dictionary. According to this method, there is no problem for the planed languages, but there is serious problem when we add a new language, we need to re-select the pivot of corresponding words, and it takes much cost.

To avoid this problem, first we establish concepts (or words for each language, and for all these concepts, we define the relation between concepts, such as equivalent, or similar, or upper, lower. Then all concepts are taken position in the concept hierarchy. In the machine translation system, for equivalent concept, the system find directly target word, but for not equivalent concept, the system searches similar or upper or lower concept , from concept hierarchy then finds target word.

There are two merits for doing this. First, merit is that, we do not need to select the pivot of corresponding words to establish the interlingua vocabulary which consists of word meaning of each language, and it reduce the cost. Second merit is that, it is easy and low cost to add a new language. The reason is mentioned above.

5 Conclusion

In designing the interlingua aimed at multilingual translation we decided to organize and represent information expressed by sentences from the four perspectives: concept of a sentence, speaker's perspective, speaker's intention, and sentence structure. By doing so, we were able to form a structure in which language representations related to tenses and aspects tending to depend on the language could be expressed in universal representations.

In CICC multilingual machine translation systems for Japanese, Chinese, Indonesian, Malay and Thai language are under development. This interlingua is used as the interface of these systems. Now, the evaluation of the performance of the interlingua and total system is undergoing.

There are two subjects in future. One is the standardization of interlingua. For fully accepting the merit of interlingua approach, it is necessary to make the standardization of interlingua. We believe that this interlingua will be the material basis of the standard interlingua. Second is the establishment of the concept hierarchy for all the concept in all languages. This kind of concept hierarchy is indispensable to establish the standard interlingua.

Acknowledgments

This research is the result of a part of the cooperation between neighboring countries on researches related to machine translation systems. We would like to thank the members of the Interlingua Committee and the EDR-CICC Technological Communication Committee for their cooperation and advice.

References

[l] Hiroshi UCHIDA, Meiying ZHU: An Interlingua for Multilingual machine Translation. 89-NL-72-9. Information Processing Society of Japan, 1989.

[2] Hiroshi UCHIDA: ATLAS II : A Machine Translation System Using Conceptual Structures as an Interlingua. Machine Translation Summit, pp. 85-92 1989.

[3] Itsuki HOSOE : An inquiry into the moaning of tense in the English verb, Shinozaki-Shorin. 1931.

[4] Nagao. Tanaka. Makino. Uchida. Ishizaki : Machine Translation Summit, Ohmsha. 1989.

[5] Japan Electronic Dictionary Research Institute. Ltd. : Concept Dictionary, TR- 27, 1990.

[6] Donna Gates, Dawn HABERLACH, etc. : Lexicons, MACHINE TRANSLATION, Vol.4. No.l . 1989.

[7] Sergei NIRENBURG, Lori LEVIN. : Knowledge Representation Support. MACHINE TRANSLATION, Vol.4. No.l 1989.

[8] Sergei NIRENBURG : Knowledge-Based Machine Translation. MACHINE TRANS-LATION. Vol. 4, No.l 1989.

[9] Tanaka, Tokunaga, etc.: A Study on Automatic Extraction of Interlingual Concepts from Bilingual Dictionaries. 89-NL-72-3, Information Processing Society of Japan, 1989

[10] Uchida, Zhu: Interlingua (Final Edition). CICC. 1993.3

[11] Masuoka. Takubo: Fundamental Japanese Grammar, Kurosio Shuppan, 1989.9