SOME LINGUISTIC ASPECTS FOR AUTOMATIC TEXT UNDERSTANDING

Yutaka Kusanagi
Institute of Literature and Linguistics
University of Tsukuba
Sakura-mura, Ibaraki 305 JAPAN

ABSTRACT

This paper proposes a system of mapping classes of syntactic structures as instruments for automatic text understanding. The system illustrated in Japanese consists of a set of verb classes and information on mapping them together with noun phrases, tense and aspect. The system, having information on direction of possible inferences between the verb classes with information on tense and aspect, is supposed to be utilized for reasoning in automatic text understanding.

I. INTRODUCTION

The purpose of this paper is to propose a system of mapping classes of syntactic structures as instruments for automatic text understanding. The system consists of a set of verb classes and information on mapping them together with noun phrases, tense and aspect, and is supposed to be utilized for inference in automatic text understanding.

The language used for illustration of the system is Japanese.

There is a tendency for non-syntactic analysers and semantic grammars in automatic text understanding. However, this proposal is motivated by the fact that syntactic structures, once analyzed and classified in terms of semantic relatedness, provide much information for understanding. This is supported by the fact that human beings use syntactically related sentences when they ask questions about texts.

The system we are proposing has the following elements:

1) Verb classes.

2) Mapping of noun phrases between or among some verb classes.

3) Direction of possible inference between the classes with information on tense and aspect.

Our experiment, in which subjects are asked to make true-false questions about certain texts, revealed that native speakers think that they understand texts by deducting sentences lexically or semantically related. For instance, a human being relates questions such as 'Did Mary go to a theater?' to a sentence in texts such as 'John took Mary to a theater.' Or, by the same sentence, he understands that 'Mary was in the theater.'

II. FEATURES OF THE JAPANESE SYNTAX

Features of Japanese syntax relevant to the discussion in this paper are presented below.

The sentence usually has case markings as postpositions to noun phrases. For instance,

1. John ga Mary ni himitsu o hanashita
'John told a secret to Mary.'

In sentence 1. postpositions ga, ni and o indicate nominative, dative and accusative, respectively.

However, postpositions do not uniquely map to deep cases. Take the following sentences for example.

2. John wa sanji ni itta.
'John went at 3 o'clock.'

3. John wa Tokyo ni itta.
'John went to Tokyo.'

4. John wa Tokyo ni sundeiru.
'John lives in Tokyo.'

Ni in the sentences 2, 3, 4 indicate time, goal and location, respectively. This is due to the verb category (3 and 4) or the class of noun phrases (2 and 3) appearing in each sentence.

Certain morpheme classes hide the casemarking, e.g.

5. John mo itta.
'John also went (somewhere).

6. Tokyo mo itta.
'Someone went to Tokyo also.'

The mo in sentence 5 and 6 means 'also'. Therefore these sentences are derived from different syntactical constructions, that is, sentences 7 and 8, respectively.

7. John ga itta.
'John went (somewhere).'

8. Tokyo ni itta.
'Someone went to Tokyo.'

Furthermore, as illustrated in sentences 5 through 6, noun phrases may be deleted freely, provided the context gives full information. In sentences 6 and 7, a noun phrase indicating the goal is missing and sentences 6 and 8 lack that indicating the subject. Finally, there are many pairs of lexically related verbs, transitive and instrasitive, indicating the same phenomenon differently

9. John ga Mary ni hon o miseta.
'John showed a book to Mary.'

10. Mary ga hon o mita.
'Mary saw a book.'

The two expressions, or viewpoints, on the same phenomenon, that is, 'John showed to Mary a book which she saw.' are related in Japanese by the verb root mi.

The system under consideration utilizes some of the above features (case marking and lexically related verbs) and in turn can be used to ease difficulties of automatic understanding, caused by some other features (case hiding, ambiguious case marking and deletion of noun phrases.)

III. VERB CLASS

The system is illustrated below with verbs related to the notion of movement. The verb classes in this category are as follows:

(1) Verb class of causality of movement(CM)

Examples: tsureteiku 'to take (a person)'
tsuretekuru 'to bring (a person)'
hakobu 'to carry'
yaru 'to give'
oshieru 'to tell'

Verbs of this class indicate that someone causes something or someone moves. How to move varies as seen later.

(2) Verb class of movement(MV)

Examples: iku 'to go'
kuru 'to come'
idousuru 'to move'

Verbs of this class indicated that something or someone moves from one place to another.

(3) Verb class of existence(EX)

Examples: iru '(animate) be'
aru '(inanimate) be'

Verbs of this class indicate the existence of something or someone.

(4) Verb class of possession(PS)

Examples:motsu 'to possess'
kau 'to keep'

Verbs of this class indicate someone's possession of something or someone.

Notice that the fundamental notion of MOVE here is much wider than the normal meaning of the word 'move'. When someone learns some idea from someone else. it is understood that an abstract notion moves from the former to the latter.

Furthermore, verbs of each class differ slightly from each other in semantic structures. But the difference is described as difference in features filling the case slot. As seen below. the difference between yaru, 'to give' and uru, 'to sell' is that the latter has 'money' as instrument, while the former does not. Incidentally, Japanese has a verb yuzuru which can be used whether the instrument is money or not.

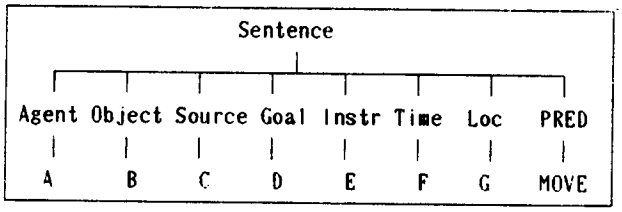

Diagram I: Semantic Structure

IV. MAPPING OF SYNTACTIC STRUCTURES

Suppose sentences of the verb of MOVE have a semantic fram roughly as illustrated in Diagram I.

The relationship among the surface

A ga	B o	C kara	D ni	E de	CM
A ga	B o	C kara	D ni	E de	MVsase
	B ga	C kara	D ni	E de	MV
	B ga	C kara	D ni	E de	CMrare
	B ga		D ni		EX
D ga	B o				PS

(sase and rare indicate causative and passive expressions respectively.)

Diagram II:Mapping of Syntactic Structures

			Obj	Goal	Source	Inst
CV	tsureteiku	take	+ani	+loc		=Agt
	mottekuru	bring - to	-ani	+loc		=Agt
		bring - for	-ani	+ani		=Agt
	hakobu	carry		+loc		=Agt
	yaru	give		+ani	=Agt	-money
	uru	sell		+hum	=Agt	+money
	oshieru	tell	+abs	+ani	=Agt	
	osowaru	learn	+abs	=Agt	+ani	
MV	iku	go		+loc		
	idousuru	move		+loc		
	tsutawaru	be conveyed	+abs		+ani	
EX	iru	be	+ani			
	aru	be	-ani			
PS	motsu	have	(-anim)			
	kau	keep	+anim			

(ani, anim, hum, abs and loc indicate animate, animal human, abstract and location, respectively)

Diagram III: Verbs and conditions for realization

syntactic structures of the verb classes discusssed above is presented in Diagram II.

Items filling the case slots in the semantic frame, or the noun phrases in surface syntactic structures, have particular conditions depending on individual verbs. Some examples of conditions are presented in Diagram III.

By these conditions, the mapping of syntactic structures presented in Diagram II is transformed to that in terms of individual verbs. Furthermore, rules of directions for reasoning presented in Diagram IV connect specific sentences.

Take the following sentence for example.

11. John ga keiki o Mary ni mottekita.
(+ani) (-ani) (+ani) (CV-past)
'John brought a cake for Mary.'

has related sentences like the following.

12. Keiki ga Mary ni itta.
'A cake went to Mary.'

13. Keiki ga Mary (no tokoro) ni aru.
'There is a cake at Mary's'

14. Mary ga keiki o motteiru.
'Mary has cake.

As far as all the rules and conditions are incorporated into the computer program, inference would be possible among sentences 11 through 14 in automatic text understanding. Furthermore, this system can also be utilized in the automatic text understanding by locating missing noun phrases and determining ambiguous grammatical cases in the sentence, finding semantically related sentences between the questions and the text, and gathering the right semantic information.

Since this system uses information on syntactic structures, it is much simpler in terms of the semantic structures than the Conceptual Dependencey Model, for instance, and the mapping among the sentence patterns semantically related much more explicit.

REFERENCE

Fillmore, C. 1968. The case for case. IN E. Back and R. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart, and Winston.

Kusanagi, Yutaka et al. to appear. Syntax and Semantics II (in Japanese). Tokyo: Asakura Shoten.

Schank, R.C., and Abelson, R.P. 1977. Scripts, plans, goals, and understanding. Hillsdale, N.J.: Lawrence Erlbaum.


```
1)  CM     <==>MVsase  (The arrow indicates the direction
    CM     <==>CMrare         for reasoning.
    CM     <==>MV           == indicates that reasoning is
    MVsase<==>MV               possible anytime, and
    MV     <==>CMrare       -- indicates that reasoning may
    MV     <==>PS               be impossible if further
2)  MV      -->EX               information on MOVEMENT is
    CV      -->EX               is provided in the context.)
    MVsase -->EX       Condition by tense and aspect
    CMrare -->PS         1) Same tense and aspect on both
    MV     -->PS               of the arrow
    CV     -->PS           Per(fect),Past-->Imp(erfect),Non-Past
    MVsase -->PS         2)Imp, Non-Past -->Non-Past
    CVrare -->PS           Past          -->Past
```

Diagram IV: Direction and condition for reasoning