A Parallel Cooperative Natural Language Processing System : *Laputa*

PURPOSE

The purpose of this study is to develop a basic software technology which integrates every part of natural language processing: morphological analysis, syntactic analysis, semantic analysis and so on, to make best use of the advantages of the Parallel Inference Machine.

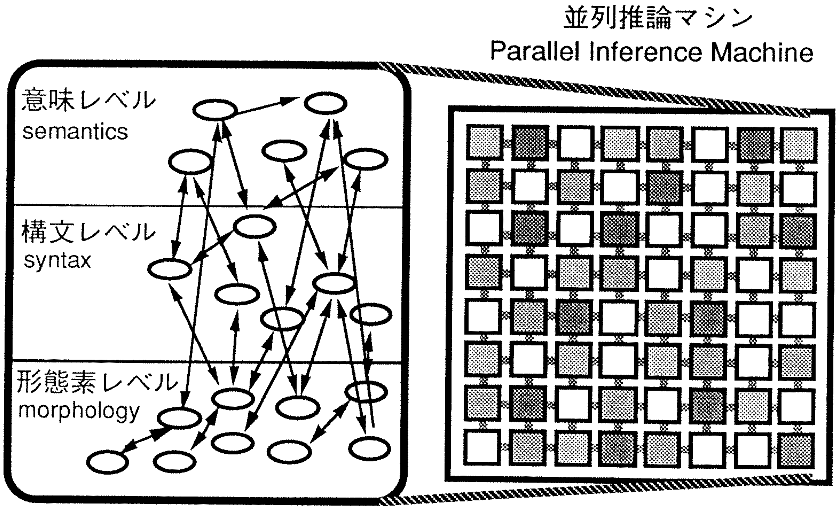
OUTLINE & FEATURES

(1) **Automatic parallel cooperation**

We propose a parallel cooperation model for natural language processing in which all the partial processing are merged together by a common processing mechanism.

(2) **Performance**

We have achieved a speed-up rate 13 utilizing 32 processors of Multi-PSI.



System Configuration

ABSTRACT

The purpose of this study is to propose a natural parallel cooperation model for natural language processing (NLP) by making use of the advantages of the Parallel Inference Machine.

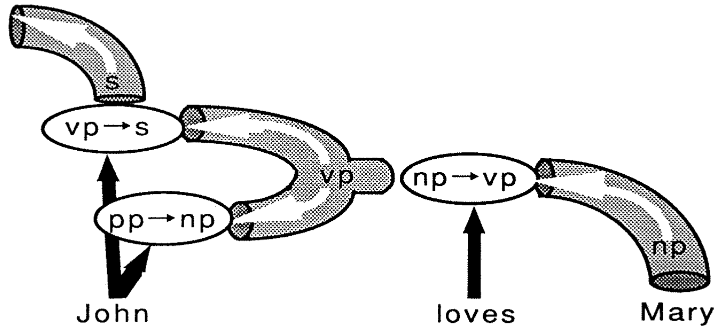
In recent years, system integration, including morphological analysis, syntactic analysis and semantic analysis, has been proposed in the NLP field.

As the basis of this proposal, it is recognized that information processing done by human has been carried out under partiality or incompleteness of information. Thus integrated NLP can adopt parallel processing because it disregards the processing direction. Therefore the integration of NLP and parallel cooperative processing can be regarded as a natural model.

However, there are few implemented systems based on this model. It is because efficient parallel cooperation requires all processes to exchange all of their information with each other, but information exchange and its control is hard to implement. One solution to this problem is to abstract the processing framework so that analysis phases such as morphological analysis, syntactic analysis are carried out by one single processing mechanism. Our processing framework utilize type inferencing with respect to record-like type structure.

APPROACH

Some efficient algorithms already exist for morphological and syntactic processing whose knowledge we should not ignore in developing a practical system, even in the case of an integrated NLP system.



Parsing based on *layered stream method*

We have found that there is a strict correspondence between our vertical type judgment and already established syntactic analysis methods. Matsumoto's parallel parser PAX performs syntactic processing in parallel through the *layered stream method*, which is an efficient mechanism for search problems based on parallel logic programming languages. Basically, PAX employs a chart parsing algorithm. Our type judgment processing formalism is quite the same as PAX, but the usage of process and data is completely opposite.

EXAMPLE OF GRAMMATICAL AND LEXICAL DESCRIPTION

The grammar rules below show how the adequacy of the semantic category of the grammatical object can vary dynamically depending on the subject.

$$\{\mathrm{np},[\mathrm{sem}=Subj]\} < (\{\mathrm{vp},VP\} \rightarrow \{\mathrm{s},VP=[\mathrm{agent}=Subj]\})$$

This grammar rule shows that the superordinate type of the type "np" is a function of type "vp → s" and that an object which is judged as the type "np" is also a function which, if applied to an object of type "vp", results in an object of type "s". In this rule, all the descriptions of "vp" are merged to those of "s", and the value of the feature "sem" of "np" is unified with the value of feature "agent" of the type "s".

Now we also show some lexicon entries to go with these rules.

eats: {np,[sem=*Obj*]} → {vp,[agent=*Ag*:{animal,[eat_obj=*Obj*]}]}
john: {np,[sem={human,*Id*,[name='John']}]}
the_tiger: {np,[sem={tiger,*Id*,[]}]}

In the lexicon the object "eats" has function type "np → vp" added with the description that says the value of the feature "agent" is a typed variable and the type of the variable is sorted "animal" and the value of the feature of "eat_obj" is unified with the value of the feature of "sem" of type "np". The rules specifying semantic categories look as follows.

{tiger,[]} < {animal,[eat_obj=*E*:{animal,[]}]}
{human,[]} < {animal,[eat_obj=*E*:{food,[]}]}

These rules mean that a tiger is an animal which eats animals and a human is an animal which eats food, respectively.

Although, under these rules of grammar, lexicon and semantic categories, 'john' and 'the_tiger' are both animals, the judgment (the_tiger,eats,john):s succeeds but (john,eats,the_tiger):s fails, because John is a human and a human is an animal which eats food but a tiger cannot be judged as food from the rules governing semantic categories.

ABSTRACT OF THE SYSTEM

The size of grammar and dictionary are as follows.

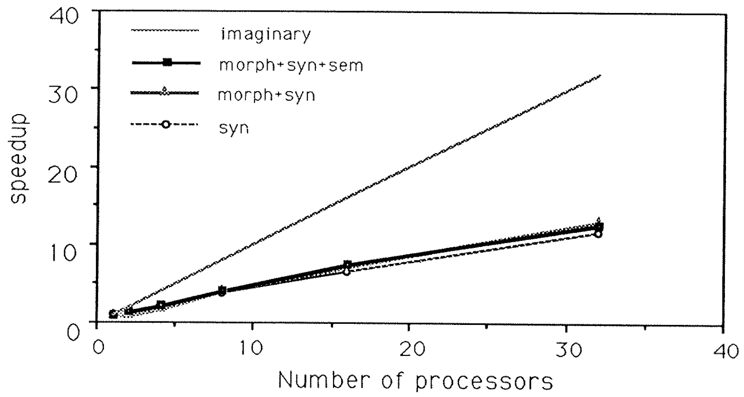
grammar rules	651
words	14,613
morphemes	8,268
concepts	770

We used the syntactic grammar and morphological grammar which were developed by Sano in ICOT 6th Laboratory. We made the conceptual system rules in accordance with the conceptual system of the Japan Electronic Dictionary Research Institute, EDR.

RESULT

As the result of our experiment, we have achieved a speed-up rate about 13 utilizing 32 processors of Multi-PSI.

Speed-up rate



OUTLINE OF DEMONSTRATION

1. Automatic parallel cooperation
2. Dynamic determination of semantic relation of subject and object
3. Parallel cooperation of morphological and syntactic analysis
4. Parallel cooperation added with semantic processing
5. Speed-up demonstration