Dynamic Programming Method for Analyzing Conjunctive Structures in Japanese

Sadao Kurohashi and Makoto Nagao
Dept. of Electrical Engineering, Kyoto University
Yoshida-honmachi, Sakyo, Kyoto, 606, Japan
kuro@kuee.kyoto-u.ac.jp

Abstract

Parsing a long sentence is very difficult, since long sentences often have conjunctions which result in ambiguities. If the conjunctive structures existing in a long sentence can be analyzed correctly, ambiguities can be reduced greatly and a sentence can be parsed in a high successful rate. Since the prior part and the posterior part of a conjunctive structure have a similar structure very often, finding two similar series of words is an essential point in solving this problem. Similarities of all pairs of words are calculated and then the two series of words which have the greatest sum of similarities are found by a technique of dynamic programming. We deal with not only conjunctive noun phrases, but also conjunctive predicative clauses created by "Renyoh chuushi-ho". We will illustrate the effectiveness of this method by the analysis of 180 long Japanese sentences.

1 Introduction

Analysis of a long Japanese sentence is one of many difficult problems which cannot be solved by the continuing efforts of many researchers and remain abandoned. It is difficult to get a proper analysis of a sentence whose length is more than fifty Japanese characters, and almost all the analyses fail for sentences composed of more than eighty characters. To clarify why it is is also very difficult because there are varieties of reasons for the failures. People sometimes say that there are so many possibilities of modifier/modifyee relations between phrases in a long sentence. But no deeper consideration has ever been given for the reasons of the analysis failure. Analysis failure here means not only that no correct analysis is included in the multiple analysis results which are caused by the intrinsic ambiguity of a sentence and also by inaccurate grammatical rules, but also that the analysis fails in the middle of the analysis process.

We have been claiming that many (more than two) linguistic components are to be seen at the same time in a sentence for proper parsing, and also that tree to tree transformation is necessary for reliable analysis of a sentence. Popular grammar rules which merge two linguistic components into one are quite insufficient to describe the delicate relationships among components in a long sentence.

Language is complex. There often happens that components which are far apart in a long sentence cooccur, or have certain relationships. Such relations may be sometimes purely semantic, but often they are grammatical or structural, although they are not definite but very subtle.

A long sentence, particularly of Japanese, contains parallel structures very often. They are either conjunctive noun phrases, or conjunctive predicative clauses. The latter is called "Renyoh chuushi-ho". They appear in an embedded sentence to modify nouns, and also are used to connect two or more sentences. This form is very often used in Japanese, and is a main cause for structural ambiguity. Many major sentential components are omitted in the posterior part of Renyoh chuushi expressions and this makes the analysis more difficult.

For the successful analysis of a long Japanese sentence, these parallel phrases and clauses, including Renyoh chuushi-ho, must be recognized correctly. This is a key point, and this must be achieved by a completely different method from the ordinary syntactic analysis methods, because they generally fail in the analysis for a long sentence.

We have introduced an assumption that these parallel phrases/clauses have a certain similarity, and have developed an algorithm which finds out a most plausible two series of words which can be considered parallel by calculating a similarity measure of two arbitrary series of words. This is realized by using the dynamic programming method. The results was exceedingly good. We achieved the score of about 80% in the detection of various types of parallel series of words in long Japanese sentences.

2 Types of Conjunctive Structures and Their Ambiguities

First, we will explain what kind of conjunctive structures (hereafter abbreviated as 'CS') appear in Japanese[1][2].

The first type is **conjunctive noun phrases**. We

Table 1: Words indicating conjunctive structures.

(a) Conjunctive noun phrases
,[comma] と も や か とか かつ だけで(は) なく および または ならびに あるいは もしくは

(b) Conjunctive predicative clauses
の+に対し(て) とか か し が ず+に だけで(は) なく けれど(も) および または ならびに あるいは もしくは

(c) Conjunctive incomplete structures
および または ならびに あるいは もしくは

'+' means succession of words. Characters in '( )' may or may not appear.

Table 2: Examples of conjunctive structures.

Conjunctive noun phrases
(i) ... 解析 (*analysis*) <u>と</u>(*and*) 生成を (*generation*) ...
(ii) ... 原言語 (*source language text*) の (*of*) 解析 (*analysis*) <u>と</u>(*and*) 相手言語 (*target language text*) の (*of*) 生成を (*generation*) ...
(iii) ... 原言語を (*source language text*) 解析する (*analyzing*) 処理 (*processing*) <u>と</u>(*and*) 相手言語を (*target language text*) 生成する (*generating*) 処理を (*processing*) ...

Conjunctive predicative clauses
(iv) ... 原言語を (*source language text*) <u>解析し</u> (*analyzing*), 相手言語を (*target language text*) 生成する (*generating*) (処理を (*processing*) ...).
(v) ... 解析 (*analysis*) では (*for*) 利用する (*use*) <u>が</u>(*but*), 生成 (*generation*) では (*for*) 利用しない (*do not use*) (という (*as*) ...).

Conjunctive incomplete structures
(vi) ... 前者 <u>を</u>(*the former*) 解析 (*analysis*) <u>に</u>(*for*), 後者 <u>を</u>(*the latter*) 生成 (*generation*) <u>に</u>(*for*) ...
(vii) ... 解析 (*analysis*) に (*for*), <u>または</u>(*and*) 生成 (*generation*) に (*for*) ...

can find these phrases by the words for conjunction listed up in Table 1(a). Each conjunctive noun sometimes has adjectival modifiers (Table 2(ii)) or clause modifiers (Table 2(iii)).

The second type is **conjunctive predicative clauses**, in which two or more predicates[1] are in a sentence forming a coordination. We can find these clauses by the **Renyoh-forms**[2] of predicates (Renyoh chuushi-ho: Table 2(iv)) or by the predicates accompanying one of the words in Table 1(b) (Table 2(v)).

The third type is CSs consisting of parts of conjunctive predicative clauses. We call this type **conjunctive incomplete structures**. We can find these structures by the correspondence of postpositional particles (Table 2(vi)) or by the words in Table 1(c) which indicate CSs explicitly (Table 2(vii)).

For all of these types, it is relatively easy to find the existence of a CS by detecting a **distinctive key bunsetsu**[3] (we call this bunsetsu **'KB'**) which accompanies these words explained above. KB lies last in the prior part of a CS, but it is difficult to determine which bunsetsu sequences on both side of the KB constitute a CS. That is, it is not easy to determine which bunsetsu to the left of a KB is the leftmost element of the prior part of a CS, and which bunsetsu to the right of a KB is the rightmost element of the posterior part of a CS. The bunsetsus between these two extreme elements constitute the **scope of the CS**. Particularly in detecting this scope of a CS, it is essential to find out the last bunsetsu in the posterior part of the CS, which corresponds to the KB. There are many candidates for it in a sentence; e.g., in a conjunctive noun phrase all nouns after a KB are the candidates. We call such a candidate bunsetsu **'CB'**. It is almost impossible to solve this problem merely by using rules based on phrase structure grammar.

[1]In addition to verbs and adjectives, assertive words (kinds of postpositions) "だ"(da), "である"(dearu), "です"(desu) and so on, which follow directly after nouns, can be predicate in Japanese.

[2]The ending forms of inflectional words which can modify verb, adjective, or assertive word are called Renyoh-form in Japanese.

[3]Bunsetsu is the smallest meaningful block consisting of an independent word (**IW**; nouns, verbs, adjectives, etc.) and accompanying words (**AW**; postpositional particles, auxiliary verbs, etc.).

3 Analysis of Conjunctive Structures

We detect the scope of CSs by using wide range of information around a KB.[4] An input sentence is first divided into bunsetsus by the conventional morphological analysis. Then we calculate similarities of all pairs of bunsetsus in a sentence, and calculate a sum of similarities between a series of bunsetsus on the left of a KB and a series of bunsetsus on the left of a CB. Of all the pairs of the two series of bunsetsus, the pair which has the greatest sum of similarities is determined as the scope of the CS. We will explain this process in detail in the following.

3.1 Similarities between Bunsetsus

An appropriate similarity value between bunsetsus is given by the following process.

- If the parts of speech of **IWs** (independent words) are equal, give <u>2 points</u> as the similarity values. Then go to the next stage and add further the following points.

  1. If IWs match exactly (by character level) each other, add <u>10 points</u> and skip the next two steps and go to the step 4. If IWs are inflected, infinitives are compared.
  2. If both IWs are nouns and they match partially by character level, add <u>the number of matching characters × 2 points</u>.

[4]We do not handle conjunctive predicative clauses created by the Renyoh-forms of predicates (Renyoh chuushi-ho) which do not accompany comma, because almost all of these predicates modify the next nearest predicate and there is no need to check the possibility of conjunction.

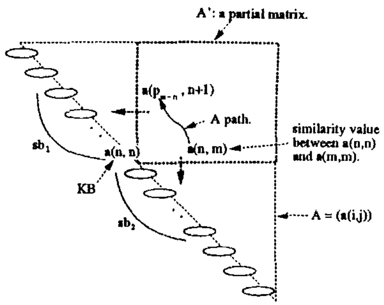


Figure 1: A path.

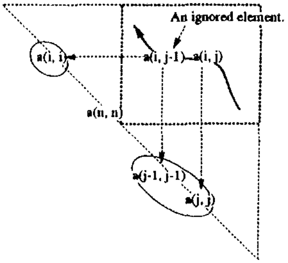


Figure 2: An ignored element.

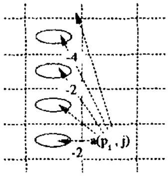


Figure 3: Penalty points.

3. Add points for semantic similarities by using the thesaurus 'Bunrui Goi Hyou' (**BGH**)[3]. BGH has the six layer abstraction hierarchy and more than 60,000 words are assigned to the leaves of it. If the most specific common layer between two IWs is the k-th layer and if k is greater than 2, add (k − 2) × 2 points. If either or both IWs are not contained in BGH, no addition is made. Matching of the generic two layers are ignored to prevent too vague matching in broader sense.
4. If some of **AWs** (accompanying words) match, add the number of matching AWs × 3 points.

Maximum sum of the similarity values which can be added by the steps 2 and 3 above is limited to 10 points.

- Although the parts of speech of IWs are not equal, give 2 points if both bunsetsus can be predicate (see footnote 1).

For example, the similarity point between "低水準言語 (*low level language*) +," and "高水準言語 (*high level language*) + と (*and*)" is calculated as 2(match of parts of speech) + 8(match of four characters: 水準言語) = 10 points. The point between "訂正 (*revision*) + し (*do*) +," and "検出 (*detection*) + する (*do*)" is 2(match of parts of speech) + 2(match by BGH) + 3(match of one AWs) = 7 points.

3.2 Similarities between Two Series of Bunsetsus

Our method detects the scope of a CS by two series of bunsetsus which have the greatest similarity. These two series of bunsetsus are searched for on a triangular matrix $A = (a(i,j))$ (Figure 1), whose diagonal element $a(i,i)$ is the i-th bunsetsu in a sentence and whose element $a(i,j)$ ($i < j$) is the similarity value between bunsetsu $a(i,i)$ and bunsetsu $a(j,j)$.

We call the rectangular matrix A' a **partial matrix**, where

$$A' = (a(i,j)) \; (0 \le i \le n; \; n+1 \le j \le l)$$

is the upper right part of a KB (Figure 1). In the following, l indicates the number of bunsetsus and $a(n,n)$ is a KB. We define a **path** as a series of elements from a non-zero element in the lowest row to an element in the leftmost column of a partial matrix (Figure 1).

$$\begin{aligned} &\text{path} ::= \\ &\quad (a(p_1, m), a(p_2, m-1), \ldots, a(p_{m-n}, n+1)), \\ &\text{where } n+1 \le m \le l, \; a(p_1, m) \ne 0, \; p_1 = n, \\ &\qquad p_i \ge p_{i+1} (1 \le i \le m-n-1). \end{aligned}$$

The starting element of a path shows the correspondence of a KB to a CB. A path has only one element from each column and extends towards the upper left.

We calculate the similarity between the series of bunsetsus on the left side of the path (sb_1 in Figure 1) and the series under the path (sb_2 in Figure 1) as a **path score** by the following four criteria:

1. Basically the score of a path is the sum of each element's points on the path. But if a part of the path is horizontal $(a(i,j), a(i,j-1))$ as shown in Figure 2, which leads the bunsetsu correspondence of one element $a(i,i)$ to two elements $a(j-1,j-1)$ and $a(j,j)$, the element's points $a(i,j-1)$ is not added to the path score.

2. Since a pair of conjunctive phrases/clauses often appear as a similar structure, it is likely that both conjunctive phrases/clauses contain nearly the same numbers of bunsetsus. Therefore, we impose penalty points on the pair of elements in the path which causes the one-to-plural bunsetsu correspondence so as to give a priority to the CS of the same size. Penalty point for

Table 3: Separating levels (SLs).

Level	Condition to Bunsetsu
5	Being the KB of a conjunctive predicative clause, or accompanying a topic-marking postpositional particle "は" and comma.
4	Accompanying a postpositional particle not creating a conjunctive noun phrase and comma, or being an adverb accompanying comma.
3	Being the Renyoh-form of a predicate which does not accompany comma, or accompanying a topic-marking postpositional particle "は".
2	Being the KB of a conjunctive noun phrase accompanying comma.
1	Accompanying a comma, or being the KB of a conjunctive noun phrase not accompanying comma.

$(a(p_i, j), a(p_{i+1}, j-1))$ is calculated by the formula (Figure 3),

$$|p_i - p_{i+1} - 1| \times 2.$$

The penalty points are subtracted from the path score.

3. Since each phrase in the CS has a certain coherency of meaning, special words which separate the meaning in a sentence often limit the scope of a CS. If a path includes such words, we impose penalty points on the path so that the possibility of including those are reduced. We define five 'separating-levels' (SLs) for bunsetsus, which express the strength of separating a sentence meaning (Table 3, cf. Table 1). If bunsetsus on the left side of the path and under it include a bunsetsu whose SL is equal to KB's SL or higher than it, we reduce the path score by

$$(\text{SL of the bunsetsu} - \text{KB's SL} + 1) \times 7.$$

However, two high SL bunsetsus corresponding to each other often exist in a CS, and those do not limit the scope of the CS. For example, topic-marking postpositional particles correspond each other in the following sentential style,

A として は (*As to* A), ...であり (*be*),
B として は (*as to* B), ...である (*be*).

Therefore, when two high SL bunsetsus correspond in a CS, that is, the path includes the element which indicates the similarity of them, and those are the 'same-type', the penalty points on them are not added to the path score. We define the same-type bunsetsus as two bunsetsus which satisfy the following two conditions.

- IWs of them are of the same part of speech, and they have the identical inflection when they are inflectional words.
- AWs of them are identical.

Table 4: Words for bonuses.

Conjunctive noun phrases	
last AW	など 等
next IW	各～ ～種類 ～つ 組 対 両方
Conjunctive predicative clauses	
last AW	ために ための という といった ようだ など 等
next IW	こと もの とき 方式 方法 手法

4. Some words frequently become the AW of the last bunsetsu in a CS or the IW following it. These words thus signal the end of the CS. Such words are shown in Table 4. Bonus points (6 points) are given to the path which indicates the CS ending with one of the words in Table 4, as that path should be preferred.

3.3 Finding the Conjunctive Structure Scope

As for each non-zero element in the lowest row in a partial matrix A' in Figure 1, we search for the best path from it which has the greatest path score by a technique of the dynamic programming. Calculation is performed column by column in the left direction from a non-zero element. For each element in a column, the best partial path including it is found by extending the partial paths from the previous column and by choosing the path with the greatest score. Then among the paths to the leftmost column, the path which has the greatest score becomes the best

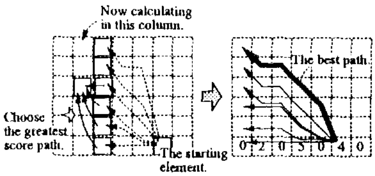


Figure 4: The best path from a element.

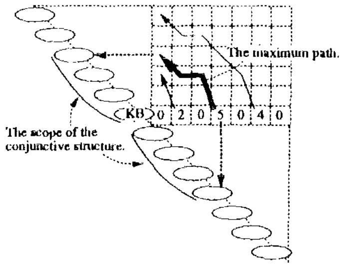


Figure 5: The maximum path specifying a conjunctive structure.

path from the non-zero element (Figure 4).

Of all the best paths from non-zero elements, the path which have the maximum path score defines the scope of the CS; i.e., the series of bunsetsus on the left side of the maximum path and the series of bunsetsus under it are conjunctive (Figure 5).

4 Experiments and Discussion

We illustrate the effectiveness of our method by the analysis of 180 Japanese sentences. 60 sentences which are longer and more complex than the average sentences are collected from each of the following three sources; Encyclopedic Dictionary of Computer Science (EDCS) published by Iwanami Publishing Co., Abstracts of papers of Japan Information Center of Science and Technology (JICST), and popular science journal, "Science", translated into Japanese (Vol.17,No.12 "Advanced Computing for Science"). Each group of 60 sentences consists of 20 sentences from 30 to 50 characters, 20 sentences from 50 to 80 characters, and 20 sentences over 80 characters.

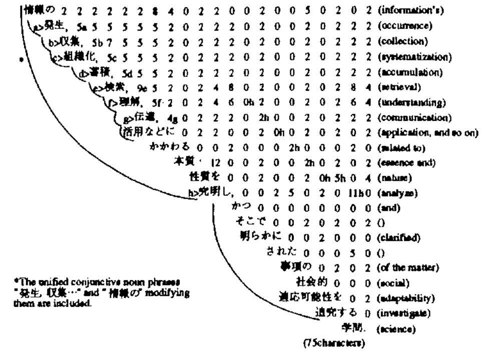
As described in the preceding sections, many factors have effects on the analysis of CSs, and it is very important to adjust the weights for each factor. The method of calculating the path score was adjusted during the experiments on 30 sentences out of 60 sentences from EDCS. Then the other 150 sentences are analyzed by these parameters. As the analyses were successful as shown in the following, this method can be regarded as properly representing the balanced weights on each factor.

This method defines where the CS ends, that is, which bunsetsu corresponds to the KB. However, as for conjunctive noun phrases containing clause modifiers or conjunctive predicative clauses, it is almost impossible to find out exactly where the CS starts, because many bunsetsus which modify right-hand bunsetsus exist in each part of the CSs and usually they do not correspond exactly. Thus it is necessary to revise the starting position of the CS obtained by this method. We treat the actual prior part of a CS as extending to bunsetsus which modify a bunsetsu in the prior part of it obtained by this method, unless they contain comma or topic-marking postpositional particle "は"(ha).

4.1 Examples of Correct Analysis

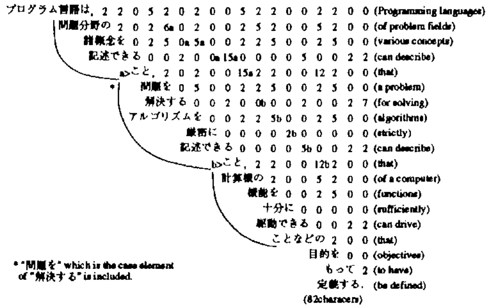
Examples of correct analysis are shown in Figure 6–8. The revisions of CS scopes are shown in notes of each figure. Chains of alphabet symbols attached to matrix elements show the maximum path concerning the KB marked by the same alphabet and '>'.

In the case of example(a) in Figure 6, the conjunctive noun phrase, in which eight nouns are conjuncted (chains of 'a', 'b', $\cdots$ 'g'), is analyzed rightly thanks to the penalty points by SLs of every comma between nouns. Thus, the CS consisting of more than two



It is a kind of science which analyzes the essence and nature related to information's occurrence, collection, systematization, accumulation, retrieval, understanding, communication, and application, and so on, and investigates social adaptability of the clarified matter.

Figure 6: An example of analyzing conjunctive structures (a).



Programming languages are defined to have objectives that they can describe various concepts of problem fields, that they can strictly describe algorithms for solving a problem, and that they can drive functions of a computer sufficiently.

Figure 7: An example of analyzing conjunctive structures (b).

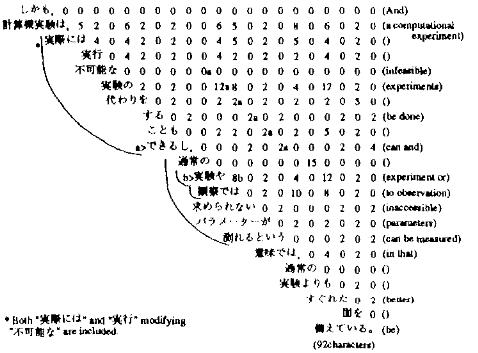
parts is expressed by the repetition of the combination of CSs consisting of two parts. In this example, also the conjunctive predicative clause is analyzed rightly (chains of 'h').

In the case of example(b) in Figure 7, the CS which consists of three noun phrases containing modifier clauses is detected as the combination of the two consecutive CSs like example(a) (chain of 'a' and 'b').

In the case of example(c) in Figure 8, the conjunctive noun phrase and the conjunctive predicative clause containing it is analyzed rightly. In this example, the successful analysis is due to the penalty points by SL of the topic-marking postpositional particle "は" in "計算機実験は (*a computational experiment*)" and "意味では (*in that*)" which are the outside of the CS and the bonus points by the AW "という" in the last bunsetsu of the CS .

Table 5: Results of experiments

Source		EDCS			JICST Abstracts			Science			Total
Length		30-50	50-80	80-149	30-50	50-80	80-144	30-50	50-80	80-139	
The conjunctive noun phrases	Success	5	9	9	7	7	5	5	5	10	62(75%)
	Failure	3	2	2	1	4	8	0	0	1	21
The conjunctive predicative clauses	Success	6	15	16	3	10	9	1	5	6	71(76%)
	Failure	0	1	2	1	5	5	2	2	5	23
The conjunctive incomplete structures	Success	0	2	1	0	0	0	0	0	0	3(100%)
	Failure	0	0	0	0	0	0	0	0	0	0



And a computational experiment is better in that infeasible experiments can be done and parameters inaccessible to experiment or observation can be measured.

Figure 8: An example of analyzing conjunctive structures (c).

4.2 Experimental Evaluation

We evaluated the analysis result of 180 Japanese sentences by hand. The results of evaluating every sentence by each CS type are shown in Table 5. If the same type CSs exist two or more in a sentence, the analysis is regarded as a success only when all of them are analyzed rightly.

There are 144 conjunctive noun phrases in 180 sentences, and 119 phrases among them are analyzed rightly. The success ratio is 83%. There are 118 conjunctive predicative clauses in 180 sentences, and 94 clauses among them are analyzed rightly. The success ratio is 80%. There are 3 pairs of the conjunctive incomplete structures, and all of them are analyzed rightly.

As shown in Table 5, the success rate for the sentences from JICST abstracts are worse than that of the sentences from other sources. The reason for the failures is that the sentences are often very ambiguous and confusing even for a human because they have too many contents in a sentence to satisfy the limitation of the document size.

4.3 Examples of Incorrect Analysis and Solutions for Them

We give examples of failure of analysis (Table 6, Figure 9), and indicate solutions for them. In Table 6, underlined parts show the KBs, 「 ... 」 shows the wrongly analyzed scope, and 『 ... 』 shows the right scope.

- It is essential in this method to define the appropriate similarity between words. Thus changing the similarity points for more detailed groups of parts of speech (e.g., nouns can be divided into numerals, proper nouns, common nouns, and action nouns which become verbs by the combination with "する (*do*)") can improve the accuracy of the analysis. For example, the example(i) in Table 6 may be analyzed rightly if the similarity points between action noun "拡張 (*extension*)" and action noun "保守 (*maintenance*)" is greater than that between action noun "拡張 (*extension*)" and common noun "困難 (*difficulty*)".
- Semantic similarities between words are currently calculated only by using BGH which do not contain technical terms. If the similarity points between technical terms can be given by thesaurus, the accuracy of the analysis will be improved. Example(ii) will be analyzed rightly if greater points are given to the similarity between "アクティブ・チャート解析法 (*Active Chart Parsing*)" and "HPSG(*Head-driven Phrase Structure Grammar*)".
- By the additional usage of relatively simple syntactic conditions, some sentences which are analyzed wrongly by this method will be analyzed rightly. For example, because Japanese modifier/modifyee relations, including the relation between a verb and its case frame elements, do not cross each other, the modifier/modifyee relations in noun phrases and predicative clauses do not spread beyond each phrase or clause, except the relation concerning the last bunsetsu of them. This condition is not satisfied by the analyzed CS in the example(iii) whose prior noun phrase contains no verb related with the case frame element "文法を (*grammar*)". By this condition it can be estimated that only "自然言語の (*natural language*) 解析と (*analysis and*)" or "解析と (*analysis and*)"

Table 6: Examples of failure of analysis.

(i) これら (*these*) 解析手法の (*of analysis methods*) 共通した (*common*) 問題として (*as problems*) 文法規則が (*grammar rules*) 大きくなった (*increasing*) 場合の (*in the case*) 「規則の (*of rules*) 『拡張や (*extension and*) 保守の (*of maintenance*)』 困難が (*difficulty*)」 上げられる (*can be thought*).
(ii) ... 日本語対話文解析部は (*Japanese dialogue analytic module*), 『「解析過程の (*of the analysis process*) 制御が (*control*) 自由な (*be free*) アクティブ・チャート解析法と (*Active Chart Parsing and*) 単一化に (*on unification*) 基づいた (*based*) 語い・統辞的な (*lexicon based*) 文法的枠組みである (*being the grammatical framework*)」 HPSGを (*HPSG*)』 採用している (*be adopted*).
(iii) 「単一の (*one*) 文法を (*grammar*) 自然言語の (*natural language*) 『解析と (*analysis and*) 生成に (*generation*)』 用いる (*using*) 双方向文法の (*of bi-directional grammar*) 研究は (*the research*), 」計算言語学の上からも (*in point of computational linguistics*), 機械翻訳や (*machine translation and*) 自然言語インタフェースといった (*such as natural language interface*) 応用面からも (*from the point of view of an application*) 重要である (*be important*). (73chs)
(iv) 実際 (*in fact*), 筆者たちは (*authors*) 『「これを (*it*) 使って (*using*), 重力相互作用が (*gravitationally interacting*) 支配する (*governing*)」 天体の (*astronomical*) 運動について (*about the motion*), 高精度で (*high-precision*) 高速の (*high-speed*) 数値計算が (*numerical computation*) できる (*can*) ディジタル・オレリーという (*called Digital Orrery*) 専用コンピューターを (*special-purpose computer*) 製作している (*create*). 』
(v) ... 『「非文に対する (*for illegal sentences*) 停止性や (*termination and*) 出力する (*outputted*) 文の (*of sentences*) あいまいさの (*of ambiguities*)」 上限について (*about the maximum*)』 保証がない (*there is no guarantee*).
(vi) ... 『表現ごとに (*for every expression*) 用意した (*prepared*) 「結合価構造中の (*in a combinative structure*) 結合価要素と (*combinative elements*) 文中の (*in a sentence*) 格要素との (*between case elements*)」』 対応を (*correspondence*) ...

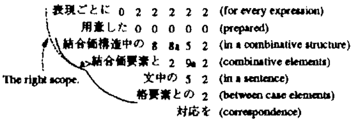


Figure 9: An example of failure of analysis.

can be the prior part of the CS. We are planning to do such a correction in the next stage of the syntactic analysis, which analyzes all modifier/modifyee relations in a sentence using the CS scopes detected by this method.

- In example(iv), the KB in the beginning part of a sentence corresponds to the last CB. That is, a short part of a sentence corresponds to the following long part. It is very difficult to analyze such an extremely unbalanced CS because this method gives a priority to similar CSs. In order to analyze example(iv) the causal relationship between "使って (*using*)" and "製作する (*create*)" will be necessary.
- Some sentences analyzed incorrectly are too subtle even for a human to find the right CSs. Example(v) cannot be analyzed rightly without expert knowledge.
- This method cannot handle the CSs in which the prior part contains some modifiers and the posterior part contains nothing corresponding to them (example(vi), Figure 9). For these structures we must think the path extending upward in a partial matrix, but it is impossible by the criteria about word similarities alone.

The CSs such as example(v) and example(vi) cannot be analyzed correctly without semantic information. However such expressions are very few in actual text.

5 Concluding Remarks

We have shown that varieties of parallel structures in Japanese sentences can be detected by the method explained in this paper. As the result, a long sentence can be reduced into a short one, and the success rate of syntactic analysis of these long sentences will become very high.

There are still some conjunctive expressions which cannot be recognized by the proposed method, and we are tempted to rely on semantic information to get proper analyses for these remaining cases. Semantic information, however, is not so reliable as syntactic information, and we have to make further efforts to find out syntactic rather than semantic relations in these difficult cases. We think that it is possible. One thing which is certain is that we have to see many more components simultaneously in a wider range of word strings of a long sentence.

References

[1] M. Nagao, J. Tsujii, N. Tanaka, M. Ishikawa (1983) Conjunctive Phrases in Scientific and Technical Papers and Their Analysis (in Japanese). *IPSJ-WG, NL-36-4*.

[2] K. Shudo, K. Yoshimura, K. Tsuda (1986) Coordinate Structures in Japanese Tehnical Sentences (in Japanese). *Trans.IPS Japan*, Vol.27, No.2, pp.183-190.

[3] The National Language Research Institute (1964) Bunrui Goi Hyou (in Japanese). Shuei Publishing.