

The Resolution of Ambiguities in Coordinate Noun Phrase Structures in Japanese:  
a corpus-based study<sup>1</sup>

日本語の名詞句並列構造における多義性の解消：コーパスに基づく研究

Nobuhiko YAMANAKA

山 中 信 彦\*

**Summary**

The resolution of ambiguities, whether actual or potential, constitutes a major part of human language understanding. In order to deal with structural ambiguities in coordinate noun phrase structures in Japanese, as in *Tokyoo no hoteru to baa no kazu* (usually interpreted as ‘the number of the hotels and bars in Tokyo’), an interactive system of three kinds of strategies, the application of which is guided by the integrated processing hypothesis, is constructed as a hypothesis out of corpus data. All of the proposed strategies are mostly based on conceptual information such as (A) correspondence of semantic categories between conjoins, (B) a ‘plurality’ concept inherent in certain head nouns, and (C) co-occurrence preferences existing between nominals and modifiers, and they are basically designed to trigger early attachments. This system of strategies applies to complex examples as well as simple ones, and corpus researches confirm its utility. In discussing complex examples, I suggest that human sentence processing is more flexible than is commonly assumed by adherents to widely accepted kinds of syntactic representations.

**Keywords** : human language understanding, structural ambiguities, integrated processing hypothesis, strategies, early attachments

**1. Introduction**

There are two outstanding aspects of ambiguity in natural language. First, it pervades in great variety our language behavior: lexical, syntactic, pragmatic, etc. In fact, it is doubtful whether we could utter a single sentence which is free from any kind of ambiguity whatsoever. Second, in spite of its variety and pervasiveness, in actual understanding of utterances we can deal with it so well that we seldom notice its occurrences. This aspect strongly suggests the

---

\* やまなか・のぶひこ  
埼玉大学教養学部教授

existence in our mental faculty of a huge amount of varied knowledge and of a special mechanism which applies this knowledge to the resolution of ambiguities.

These two aspects can be nicely illustrated by an example from Lytinen (1987: 305):

- (1) The stock cars raced by the spectators crowded into the stands at over 200 mph on the track at Indy.

Let us limit our attention to syntactic ambiguity for a moment. Without the use of (broadly construed) semantic knowledge, this sentence, Lytinen claims, requires us to consider 156 potential interpretations due to what is called a combinatorial explosion in the number of syntactic ambiguities. Let us just look at a few of them: the verbs *raced* and *crowded* could both be either past active or past participle; *at over 200 mph* could be attached to *cars*, *raced*, *spectators*, *crowded* or *stands*.

As its title shows, this paper focuses on the second aspect of ambiguity, which is more important from the standpoint of natural language understanding, namely, on how we recover the unambiguous message underlying a structurally ambiguous expression.

Traditionally this problem has been tackled by two approaches: psychological experiments and computer simulations. There is a third one, which I call a corpus-based approach.<sup>2</sup> The idea is that, given many examples involving ambiguities from a corpus as inputs and specific analyses by a native speaker as outputs, if it is possible to construct a cognitive model which predicts the outputs from the inputs, then it can be said to be a valid model on the process of disambiguation by the native speaker; in other words, that model would be, as it were, the contents of a black box. To put it more concretely, if you find in a corpus such examples as “the existence of the sun and the earth”, “the effect of the operation and medication”, and you also find that you are more likely to interpret them as “the existence of [the sun and the earth]”, “the effect of [the operation and medication]”, respectively, rather than “[the existence of the sun] and the earth”, “[the effect of the operation] and medication” even out of context, then you might propose a hypothesis that ambiguities involving coordinate noun phrase structures should be resolved in such a way that nouns of a similar semantic category correspond to each other. You will then proceed to see how far that hypothesis can account for relevant examples, and revise it if necessary. Further, if you postulate, test and revise some other hypotheses in a similar manner, until you can account for a great majority of the data by the system of the hypotheses in question, then that system is considered to be a good starting point which will eventually form a part of a cognitive model of disambiguating processes. That is the approach which I take in this paper. I do not claim that my approach is methodologically better than other approaches; rather it is complementary to an experimental one. While the latter can observe actual processes of disambiguation minutely, it is limited by the homogeneity and artificiality of used materials. A corpus-based approach, on the other hand, has the advantage of being able to study both complex examples and unexpected yet plausible examples, which

have been actually produced and which are not easily hit upon in an experimental approach.

One reason why I choose coordinate noun phrase structures as the object of analysis is related to my approach: there are so many instances of them in a corpus. Another reason is more important and is related to the combinatorial explosion mentioned above: as will be seen later, coordinate noun phrase structures can be quite long and complex, and the longer and more complex they are, the more potential structural ambiguities they produce (Nagao, 1996: 189). Yet another reason is that, despite their frequent occurrence, they have received relatively little attention from psycholinguists compared with other types of structural ambiguities.

## 2. Cognitive models

To start with, I should make clear that this paper does not propose a “model” in the usual sense but some hypotheses concerning a system of strategies for ambiguity resolution, justified with reference to the corpus data. Nonetheless, I would like to summarize some of the different theoretical positions on the modeling of human language understanding, and briefly explain how my position stands with respect to a few sets of distinction. The first, and probably the most important distinction to be made is the one between modular and interactive modelings (Mahesh et al., 1999: 29). The modular view claims that language understanding proceeds, with different linguistic components doing their jobs independently. On this view, syntactic parsing and semantic interpretation are separate processes. Most modular models are syntax-first, e.g. the garden-path model (Frazier, 1987). The interactive view claims that language understanding involves immediately integrating different types of information, such as syntactic rules, word meanings, world knowledge, etc.; in other words, humans integrate syntactic analysis and semantic interpretation. This position is called the integrated processing hypothesis by both Schank and Birnbaum (1984: 212) on the one hand and Mahesh et al. (1999: 29) on the other, although there are substantial differences between their conceptions.

Pickering et al. (2000) introduce two sets of distinction, namely, restricted versus unrestricted accounts on the one hand and parallel versus serial accounts on the other. As to the former distinction, restricted accounts propose that there is some stage during which the parser ignores some of the potentially relevant sources of information, whereas unrestricted accounts propose that all sources of information can be used in initial parsing decisions. Although Pickering et al. state that the distinction between restricted and unrestricted accounts is very general, it seems almost to coincide with the distinction between modular and interactive views just observed, considering their characterization of restricted accounts given elsewhere (Pickering et al. 2000: 12).

As for the serial versus parallel distinction, a serial account only considers one analysis at a

time, while a parallel account considers multiple analyses at the same time. Parallel accounts are further classified into momentary-parallel models and extended-parallel models, depending on whether evaluation between alternative analyses is immediate or long-lasting.

According to Pickering et al., there is a great deal of overlap between parallel and unrestricted accounts on the one hand and between serial and restricted accounts on the other, although the unrestricted race model (Van Gompel et al., 2000) is a serial but unrestricted account.

My position is not quite compatible with any of the existing models. That is not to say that I am proposing something quite different from all the others but only that it does not belong to any particular camp. It is certainly not compatible with a modular model, because I do not assume a separate syntactic parser which first computes syntactic structures and then pass them to semantics for evaluation. My position is closest, in spirit but perhaps not in detail, to that of Schank and Birnbaum (1984) and that of Lytinen (1987) in the sense that it is very much meaning-oriented. Specifically, the system of strategies I propose depends mostly on conceptual information, such as semantic categories of words, co-occurrence preferences, and discourse context. Although it is incompatible with a restricted model in the sense of being modular, it may be somewhat misleading to assign it to a unrestricted model, because unlike so called constraint-based models of which MacDonald et al. (1994) is a representative, it does not, in the present version at least, much take into consideration syntactic preferences or frequency information associated with individual lexical items. Certainly I would not go so far as to assert that these sources of information have no part to play in disambiguation process in general, but as far as the specific structures in Japanese discussed in this paper are concerned, semantic information plus simple syntax seems to make fairly good predictions.<sup>3</sup>

In regard to serial versus parallel distinction, my system of strategies is basically of a serial nature in that it tries to select the most relevant analysis as soon as the information necessary to do so is available to it. When there is doubt as to the applicability of a strategy, the system must suspend the application of it and look forward, which may in effect correspond to momentary parallelism. I will address this problem again in Section 4, where disambiguation strategies are discussed. At the very least, the system is incompatible with the standard extended-parallel model, where multiple analyses dynamically compete with each other in the ranking process (Pickering et al., 2000: 16), and it has much more in common with the unrestricted race model mentioned above.

In sum, my system of strategies is semantics-oriented, interactive, and basically serial.

### **3. The object of analysis**

The problem to be addressed is the disambiguation of coordinate noun phrase structures in Japanese. As mentioned before, it is widely recognized that coordinate noun phrase structures

contribute to potential structural ambiguities in natural language, as exemplified by the famous "old men and women". This potentiality seems to be dealt with in similar but not quite identical ways across languages. For instance, Japanese depends far less on grammatical information than English, as the following suggests:

- (2) *Kare no tuma to kodomo*  
 He GEN wife and child  
 'His wife and child'

In English, the fact that *child* is a countable noun in singular form makes it almost obligatory for it to be modified by *his*,<sup>4</sup> whereas in Japanese the absence of such a grammatical constraint makes the corresponding modification only probable. In other words, it seems easier in Japanese than in English to find examples involving potential structural ambiguities where the clues are exclusively of a semantic or pragmatic nature. This observation is the very basis on which to construct a system of disambiguation strategies given in Section 4.

In order to prevent too much complication, the data base for the hypothesis formation must be limited to a certain range of structures. The formal properties of such structures can be formulated as follows:

- (3)  $\acute{W}_1 \acute{W}_2 \dots \acute{W}_m X \text{ Co } Y \text{ no } Z_1 \text{ no } Z_2 \text{ no } \dots Z_n$ ,  
 where  $\acute{W}_i$  ( $1 \leq i \leq m$ ) is a modifier; X, Y and  $Z_j$  ( $1 \leq j \leq n$ ) are nouns; Co stands for a coordinator  
 (Huddleston and Pullum, 2002: 1275).

In (3), modifiers include adjectives, verbs, prenominal genitives ("noun *no*"), relative clauses; nouns include simple nouns, nouns with one or more prefixes and/or one or more suffixes, and compound nouns with no formal internal boundary (a hyphen or a point); in Japanese Co is realized by a conjunction (*oyobi* 'and', *narabini* 'and', *matawa* 'or', etc.) or a particle (*to* 'and', *ya* 'and', *ka* 'or', etc., excluding the case particle *to*), but here it is extended to include a simple comma.

Let us illustrate this formula with a simple example and point out the basic ambiguities involved in it:

- (4) *Tositotta otoko to onna no kekkon*  
 old man and woman GEN marriage  
 $\acute{W}_m \quad X \quad \text{Co} \quad Y \quad \text{no} \quad Z_1$

First of all,  $\acute{W}_m$  (*tositotta*) can modify either X (*otoko*) alone or both X and Y (*onna*). I will indicate these interpretations respectively as '[*tositotta otoko*] *to onna*' and '*tositotta* [*otoko to onna*]'. Secondly,  $Z_1$  (*kekkon*) can be related to either Y alone or both X and Y.<sup>5</sup> In the latter case, there is a further distinction between combinatory and segregatory interpretation (see Subsection 4.1 for detail), that is, supposing that *otoko* and *onna* refer to a particular man and woman, respectively, whether they jointly get married or they each get a different spouse.<sup>6</sup>

Let us use the following notations to indicate these distinctions:

'otoko to [onna no kekkon]'

'[otoko to onna] no kekkon' (separate marriages)

'<otoko to onna> no kekkon' (a joint marriage)

By way of a summary of the above distinctions, below are shown possible interpretations of (4) as a whole:

(5)

(a) '[tositotta otoko] to [onna no kekkon]' 'an old man and the marriage of a woman'

(b) '[[tositotta otoko] to onna] no kekkon' 'separate marriages of an old man and a woman'

(c) '<[tositotta otoko] to onna> no kekkon' 'a joint marriage of an old man and a woman'

(d) '[tositotta [otoko to onna]] no kekkon' 'separate marriages of an old man and an old woman'

(e) '<tositotta [otoko to onna]> no kekkon' 'a joint marriage of an old man and an old woman'

It must be emphasized here that what is involved in (4) (and (3), of which it is a simple subtype) is inherent *structural* ambiguities no matter what semantic features the words ( $\acute{W}_m$ , X, Y, and  $Z_1$ ) consist of. In other words, it is an idiosyncrasy of (4) that it has five more or less plausible interpretations given in (5); depending on the specific words, the structure " $\acute{W}_m$  X to Y no  $Z_1$ " may have only one plausible interpretation, but this does not make it *structurally* less ambiguous than (4).

Now it is time to describe the corpora used in this study and state the criteria for collecting examples. The main corpora are four books, the bibliographical information on which is as follows, with the number of examples found in each of them indicated in parentheses at the end:

Umesao, Tadao, 1969. *Titeki-seisan no Gizyutu*. Iwanami-syoten, Tokyo, (95).

Noguti, Yukio, 1993. *"Tyoo" Seirihoo*. Tyuuoo-kooronsya, Tokyo, (96).

Endoo, Syuusaku, 1978. *Guutara-mandansyuu*. Kadokawa-syoten, Tokyo, (59).

Sawaki, Kootaroo, 1989. *Baabon-sutoriito*. Sintyoosya, Tokyo, (50).

The former two are books on academic know-how and the latter two are essays. I chose these books as corpora for three reasons: first, they were all written within thirty years back from the outset of this study, which are not long enough for Japanese usage of coordinate noun phrase structures to have changed; second, they were all written by famous writers and some of them were best sellers at the time; third, they are well balanced in terms of formal vs. informal styles. In this paper, a cited example has the source indicated by the author's initial and the page number. An additional corpus, which was originally used in a previous study, Yamanaka (1999), is the following (I thank the late Prof. Tanaka for permitting me to make use of it):

Tanaka, Hozumi, Ogino, Takano, Ogino, Tunao, 1980. *Sinpen-nihongo-zenhinsiretu-syuusei*. Densi-gizyutu-soogoo-kenkyuuzyo-suiron-kikoo-kenkyuusitu, (219).

This corpus consists of nearly four hundred thousands separate examples, each taking one line. The examples each have a key string of four words, by the parts of speech of which they are sorted. Each example has a code consisting of a capital plus a number of six figures such as

(C155458). It corresponds to the code of a newspaper article from which the example was taken and which is included in the original texts attached to the corpus. The examples can be examined in wider contexts by looking at the original texts. In this paper, a cited example has the source indicated by its original code.

As for the criteria for sampling, the main corpora, excluding imprints, tables of contents, commentaries by critics, and indexes are searched for all the examples which fit the formula (3) (but see restrictions beginning in the next paragraph). The newspaper corpus (pp. 234-245) was searched for all the examples which fit a special type of the formula (6): " $W_1$  no  $W_2$  no..... $W_m$  no X to Y no  $Z_1$  no  $Z_2$  no..... $Z_n$ " (see formula (6) below for the notations).

The following are major patterns of examples which in the present study are not taken into account in the evaluation of the system of strategies postulated below (but a few of which are cited to illustrate a point), the reason for the exclusion being given after a colon. The purpose is to avoid their inherent context-dependence or biasing effects.

Examples in which a modifier is or contains a referring expression beginning with a demonstrative *kotosola*, e.g. *kono sora to kumo* 'this sky and cloud' (E: 41); *sono doogu no koozoo ya seinoo* (gl. 'the instrument GEN structure and function') 'the structure and function of the instrument' (U: 62): the interpretation of such an expression depends mostly on the context.

Examples in which the particle *nado* 'etc.' is present: as reported by Nagao et al. (1983), *nado* gives a formal clue as to the interpretation of coordinate noun phrase structures, namely, it suggests that the noun or noun phrase standing between the coordinator and *nado* is the last of the conjuncts, as in *kikai ya koozyoo nado no baai* (gl. 'machine and factory etc. GEN case') 'the case of machines, factories, etc.' (N: 213).

Examples in which two coordinators are present, either separately as in *isseiki-mae no bunsyoo to konniti no bunsyoo to* (gl. 'a century ago GEN text and present-day GEN text and') 'texts a century ago and present-day texts' (U: 119) or in succession as in *kari no keturon, matawa kasetu* (gl. 'tentative GEN conclusion, or hypothesis') 'a tentative conclusion or a hypothesis' (N: 171): as for the former type, Nagao et al. (1983) report that the second coordinator is a formal clue in the same sense as in the case of *nado* just mentioned; as for the latter, they report that the modifier is most unlikely to relate to the second conjunct.

#### 4. Disambiguation strategies

First, a word as to methodology is in order. I assume that there are a number of strategies which we depend on while processing ambiguous structures, independently of particular contexts in which they are used; interpretations predicted by the strategies may be endorsed or contradicted by the contexts. This assumption might appear to run counter to the spirit of the integrated processing hypothesis stated earlier. For, as an utterance naturally occurs in context and its understanding, according to the hypothesis, proceeds with all the contextual

information in view, it might well seem unrealistic or even unnecessary to think of strategies to be applied to an utterance out of context. However, there are exceptional cases, such as a title of a book, a newspaper headline, etc. which do have to be dealt with without recourse to context or, if you prefer, based on the most neutral context that we can bring to bear. Further, even with normal cases in which the structure in question is used in some context, it is most unlikely that we can do nothing about it if the structure is artificially isolated from the context, as Sperber and Wilson (1995: 185) suggest: based on encyclopaedic information about the objects and events referred to, we can more or less determine in which way it is to be interpreted.

With regard to English, a number of principles and strategies have been proposed as those used by humans in parsing input sentences in general and resolving structural ambiguities in particular. For example, Kimball (1973)'s seven principles; the minimal attachment and late closure strategies (Frazier, 1987); the principle of lexical preference, the principle of syntactic preference, the principle of final arguments, and the principle of invoked attachment (Ford et al., 1982); the principle of a priori plausibility, the principle of referential success, and the principle of parsimony (Crain and Steedman, 1985); statistically sensitive perceptual strategies (Bever et al., 1998). Specifically for Japanese, few original proposals seem to have been made. For example, the tentative attachment strategy (Mazuka and Itoh, 1995); the principle of noun phrases with *no* ("*no*" *meisi-ku no gensoku*) (Kess and Nisimitu, 1989).

For resolving ambiguities in coordinate noun phrase structures in Japanese, I will propose below a number of strategies,<sup>7</sup> which are largely of a semantic or pragmatic nature. For ease of exposition, they are introduced in three subsections: first, referring to previous studies I will state motivations for them, and illustrate them with short examples; second, I will consider them in more detail, looking at longer examples and pointing out various ramifications; third, I will consider their relationships (ordering and ranking) by analyzing complex examples. The purpose of the last subsection is to present the strategies as an interactive system rather than unrelated pieces and demonstrate its utility.

#### 4.1. Essentials

In illustrating the strategies, I will confine myself to a subtype of the structures under investigation, which has the following form:

$$(6) W_1 \text{ no } W_2 \text{ no} \dots W_m \text{ no } X \text{ Co } Y \text{ no } Z_1 \text{ no } Z_2 \text{ no} \dots Z_n,$$

This is identical to (3) except that all the modifiers preceding X are prenominal genitives "noun *no*"; otherwise the notations mean the same as in (3). The restriction to this particular subtype is motivated only by the ease with which to illustrate the strategies; they apply equally to the structures in general.

Before discussing the strategies directly related to coordinated noun phrase structures, I should point out and discuss another type of structural ambiguity found in this formula,



because it will play a major role in Subsection 4.3. That is, "W<sub>1</sub> no W<sub>2</sub> no.....W<sub>m</sub>" is itself ambiguous in that "W<sub>1</sub> no" can modify either the adjacent noun "W<sub>2</sub>" alone, the "W<sub>2</sub> no W<sub>3</sub>" as a whole, or a noun phrase farther to the right. This problem is discussed in Yamanaka (1988) and a set of strategies is proposed for its solution. The following version is slightly adapted from the original.

(7)

- (i) If "W<sub>1</sub> no W<sub>2</sub>" is a normal co-occurrence,<sup>8</sup> presume that "W<sub>1</sub> no" modifies W<sub>2</sub>, in which case "W<sub>1</sub> no W<sub>2</sub> no W<sub>3</sub>" is to be interpreted as '[W<sub>1</sub> no W<sub>2</sub>] no W<sub>3</sub>' in the normal course of events.
- (ii) If "W<sub>1</sub> no W<sub>2</sub>" does not make sense, presume that "W<sub>1</sub> no" does not modify W<sub>2</sub>; if "W<sub>1</sub> no W<sub>2</sub>" makes sense but is not a very natural collocation, suspend a decision as to whether "W<sub>1</sub> no" modifies W<sub>2</sub>. In either case, wait for the next noun W<sub>3</sub> to appear. When it appears, apply either (iii) or (iv).
- (iii) If both "W<sub>1</sub> no W<sub>3</sub>" and "W<sub>2</sub> no W<sub>3</sub>" are normal co-occurrences, presume that "W<sub>1</sub> no" modifies "W<sub>2</sub> no W<sub>3</sub>" as a whole, in which case "W<sub>1</sub> no W<sub>2</sub> no W<sub>3</sub>" is to be interpreted as 'W<sub>1</sub> no [W<sub>2</sub> no W<sub>3</sub>]' in the normal course of events.
- (iv) If it is not the case that both "W<sub>1</sub> no W<sub>3</sub>" and "W<sub>2</sub> no W<sub>3</sub>" make sense, presume that "W<sub>1</sub> no" does not modify "W<sub>2</sub> no W<sub>3</sub>" as a whole; if they both make sense but are not both very natural collocations, suspend a decision as to whether "W<sub>1</sub> no" modifies "W<sub>2</sub> no W<sub>3</sub>". In either case, wait for the next noun W<sub>4</sub> to appear. When it appears, repeat analogous procedures until a plausible interpretation is arrived at.

Let us illustrate these strategies by simple examples from Yamanaka (1988).

(8) *Hoteru no genkan no tobira*

Hotel GEN entrance GEN door

'The door of the entrance to the hotel'

In (8) "W<sub>1</sub> no W<sub>2</sub>" (*hoteru no genkan* 'the entrance to the hotel') is a perfectly normal co-occurrence. At this point, *hoteru no* is judged, according to (i), to modify *genkan*, with the result that '[*hoteru no genkan*] no tobira' is produced as a default interpretation.<sup>9</sup>

(9) *Watasi no kyoo no kibun*

I GEN today GEN mood

'My mood today'

In (9) "W<sub>1</sub> no W<sub>2</sub>" (*watasi no kyoo* 'my today') does not make very good sense (a forced interpretation may be 'my today as contrasted with your today'). At this point, *watasi no* is judged, according to (ii), not to modify *kyoo*<sup>10</sup>. On the other hand, "W<sub>1</sub> no W<sub>3</sub>" (*watasi no kibun* 'my mood') and "W<sub>2</sub> no W<sub>3</sub>" (*kyoo no kibun* 'today's mood') are both perfectly normal co-occurrences in Japanese. This, according to (iii), leads to the interpretation of (9) as '*watasi no [kyoo no kibun]*'

(10) *Nengazyoo no hebi no e*

New Year's card GEN snake GEN drawing

'The drawing of a snake on a New Year's card'

In (10) it would not be difficult to interpret " $W_1$  no  $W_2$ " (*nengazyoo no hebi*) as something like 'a snake drawn on a New Year's card'. However, *nengazyoo no hebi* is not a very natural collocation in Japanese. Therefore, according to (ii), the decision as to whether *nengazyoo no* modifies *hebi* or not is suspended. On the other hand, " $W_1$  no  $W_3$ " (*nengazyoo no e* 'the drawing on a New Year's card') and " $W_2$  no  $W_3$ " (*hebi no e* 'a drawing of a snake') are both perfectly normal co-occurrences. This, according to (iii), leads to the interpretation of (10) as '*nengazyoo no* [*hebi no e*]'.

A few comments on the above strategies will be in order. First, Strategy (i) is essentially the same as the one variously called "the preference of adjacency" (Beaugrande and Dressler, 1981: 41), "a syntactic recency preference" (Lytinen, 1987: 315), "recency preference" (Gibson and Loomis, 1994; Gibson et al., 1996), whose status seems fairly well established, at least in English. The late closure strategy (Frazier, 1987) also covers the strategy in question. There is reason to believe that it plays a more important role in Japanese than in English: in English, where a modifier often follows the head, a human parser may have already in memory some candidate attachment sites for a modifier when the recency preference is considered, whereas in Japanese, where a modifier always precedes the head, a human parser has in view only the foremost one of the potential candidates (a fact pointed out by Kamide and Mitchell (1997: 249) and by Miyamoto et al. (1999: 665))<sup>11</sup> when the preference is considered, so that s/he cannot but try attaching a modifier to this foremost candidate.

It is worth noting, however, that Kaneko (1987), based on his psychological experiment involving sentences of this type of structural ambiguity in Japanese, proposes the principle of late attachment ("*osoi huka*" no *gensoku*), which he claims is complementary to that of local attachment (Frazier and Fodor, 1978), the latter largely overlapping with the late closure strategy. His principle of late attachment is to delay the attachment of a constituent while processing the string of words which follows it, and is motivated by the result of his experiment that, for example, in *utokusii kanozyo no utago*e (gl. 'beautiful she GEN singing') 90 percent of the responses took *utokusii* to modify *utago*e rather than *kanozyo*, i.e. '*utokusii* [*kanozyo no utago*e]' 'her beautiful singing'.

Kamide and Mitchell (1997) report, on the other hand, that their on-line self-paced reading experiment suggests that the relative clause in Japanese is initially attached low, as soon as the parser encounters the first potential attachment site (NP-Low), but that by the end of the sentence the parser switches over to favoring high attachment, which is consistent with the results of their off-line questionnaire study. On the whole, studies have shown mixed results as to which noun is preferred as an attachment site, as Hirose (2006: 269) summarizes.

I should emphasize that my position is basically in favor of the early attachment account and

that not only Strategies (7) but also Strategies (A)-(C) discussed below are designed to trigger early attachments.

Second, Miyamoto et al. (1999: 666ff.) would consider the set of strategies described in (7) to be essentially along the line of *lexicalized-parsing* as opposed to *predictive-parsing*. Based on their on-line self-paced reading experiment, they argue for a predictive-parsing model. It should be noted, however, that their conception of postpositions, which they assume are major sources for information predicting a next noun to be processed, differs significantly from mine. To take *shigemi no yoko no hito no ushiro no jitensha* (gl. 'bush GEN side GEN person GEN back GEN bicycle') 'the bicycle behind the person beside the bush' for example, they treat *yoko no* and *ushiro no* as complex locative postpositions (ibid.: 669ff.), whereas I would regard the whole phrase simply as five nouns connected by *nos*.

At this point I want to make clear in what sense I described my system of strategies as being basically serial in Section 2. Van Gompel et al. (2000: 624) illustrate their model, which is unrestricted and serial, with the following example:

- (11) While the guests were eating plates were brought in.

Since the verb *to eat* has a more frequent transitive than intransitive use, the transitive analysis is initially preferred in (11). This analysis is ruled out by the semantic implausibility of *eating plates*. However, because syntactic analysis is a prerequisite for semantic analysis according to their model, the processor first needs to construct the direct object analysis before it can employ plausibility information. Thus, semantic plausibility information cannot prevent a garden path occurring, with the result that reanalysis is predicted at *plates*.

If I were to give the same kind of description of the comprehension process of (9), it would be as follows:

First, when *kyoo* is encountered, the analysis that it is modified by *watasi no* is preferred by virtue of adjacency, which has a statistical basis. Then, semantic plausibility information comes into play, canceling the modification and predicting reanalysis.

Intuitively, neither a garden path nor reanalysis seems to be involved in the processing of (9); if anything, a garden path would seem to occur if the phrase ended without *no kibun*: syntactically *watasi no* would have to modify *kyoo* since it had no other noun to modify, although semantically it would result in an anomalous co-occurrence. This situation would naturally follow from the hypothesis that *watasi no* is judged from the first not to modify *kyoo* on the grounds of semantic incongruity.

In my system, then, although I chose to call it a serial one, the decision as to modification is triggered by semantic plausibility information rather than endorsed by it. When the decision is suspended, as in (10), the system may be said to be showing momentary parallelism. What has been said so far also applies to the strategies considered below.

Returning now to strategies directly related to coordinate noun phrase structures, I propose

basically three kinds.

(A) The strategies based on corresponding semantic categories

It is pointed out by Isiwata (1965) that there is a strong tendency for coordinated nouns in context to be similar in semantic categories. This fact is utilized in machine-oriented natural language processing for resolving ambiguities in coordinate noun phrase structures (Isiwata, 1965, 1968; Nagao, 1996). Based on this idea, I propose the following generalized strategy which is specifically designed for language understanding by humans:

Strategy (A-1)

When X and Co are encountered, search in order of input for a noun or noun phrase which is likely to correspond to X semantically.

Let us begin by a very simple example (the corresponding parts are underlined).

(12) Papa to mama no mae de (A003677)

Dad and Mom GEN front LOC

'in front of Dad and Mom'

Since *papa* and *mama* belong to the same semantic category of parents, they are immediately recognized as correspondents. I will present a number of simple cases such as (12) and a few of somewhat more complex cases in the next subsection.

There is a productive pattern of examples which suggests that, while engaged in Strategy (A-1), we are also guided by this strategy:

Strategy (A-2)

When Y is encountered, search for a noun or noun phrase preceding Co and thus in your short-term memory which is likely to correspond to Y semantically.

(13) Tobasu no koobu to toden no zenbu (A024860)

Metropolitan bus GEN back and Metropolitan tramcar GEN front

'the back of the Metropolitan bus and the front of the Metropolitan tramcar'

(14) Honkon no svooboosi ya Tai no keisatukan (S: 50)

Hong Kong GEN fireman and Thailand GEN police officer

'firemen in Hong Kong and police officers in Thailand'

(15) Mukoo sankagetu no yohoo to kankooki yohoo (C154652)<sup>12</sup>

Next three months GEN forecast and cold season forecast

'the forecast for next three months and the forecast for cold seasons'

In (13) and (14), not only  $Z_1$  corresponds to X, but also Y corresponds to  $W_1$ . In (15), only the second constituent of Y (*yohoo*) corresponds to X and the first constituent (*kankooki*) corresponds to  $W_1$ .

(B) The strategy based on binding

Quirk et al. (1985: 958) distinguish between combinatory and segregatory coordination (which Huddleston and Pullum (2002: 1281) refer to as joint and distributive coordination,

respectively):

Combinatory coordination is associated with nonelliptical interpretation, and segregatory with ellipsis or quasi-ellipsis. In the combinatory interpretation.....the conjoints act semantically 'in combination' with respect to the rest of the sentence.

For example, (16), a case of segregatory coordination, may be expanded to (17), whereas (18) in its usual (but not necessary) combinatory interpretation, is not paraphrasable by (19).

- (16) They sell manual and electric typewriters.
- (17) They sell manual typewriters and electric typewriters.
- (18) They made salmon and cucumber sandwiches.
- (19) They made salmon sandwiches and cucumber sandwiches.

This distinction is referred to by Mizutani and Tanaka (1972: 22) as the one between *tabane* 'binding' and *narabe* 'juxtaposing'. In the pattern "X to Y no Z<sub>1</sub>", the presence of binding, i.e., combinatory coordination is known in most cases by the inherent semantic feature of Z<sub>1</sub>.

- (20) *Bi to sinkoo no aida ni* (J177464)

Beauty and faith GEN interval LOC  
'between beauty and faith'

- (21) *Kosuto to kooka no hikaku* (N: 19)

cost and effect GEN comparison  
'the comparison between costs and effects'

- (22) *Ama to puro no ryoohoo* (S: 139)

Amateur and professional GEN both  
'Both amateurs and professionals'

- (23) *Kobayasi to Huzisita no hutari* (A050296)

Kobayasi and Huzisita GEN two persons  
'both Kobayasi and Huzisita'

The fact that *aida*, *hikaku*, *ryoohoo*, and *hutari* represent a concept which presupposes plural entities as participants in order to be complete in meaning necessitates combinatory interpretation (which, as stated earlier, is indicated by '< >', as in '<*ama to puro*> *no ryoohoo*').<sup>13</sup>

- (C) The strategy based on co-occurrence preferences

Even if the strategy based on corresponding semantic categories (A-1) successfully applies, it is not enough to guarantee a correct interpretation. Co-occurrence preferences must also be considered (for the use of the term *preferences* rather than *restrictions*, see Cruse (2004: 219)). By co-occurrence preferences I mean (broadly construed) semantic conditions of some sort which have to be satisfied in order for lexical units in a syntagmatic relation (here a modifier-head relation) to go together naturally. In other words, non-satisfaction of co-occurrence preferences gives rise to various types of anomaly (see the next subsection).

Some studies point out, albeit informally, the role played by co-occurrence preferences in the

resolution of structural ambiguities, for example, in terms of knowledge of the world (Beaugrande and Dressler, 1981: 41), in terms of co-occurrence restrictions holding between the elements (Dik, 1968: 229) or in terms of idiosyncratic features of lexical elements (Kooij, 1971: 112). More explicitly, Wu and Furugori (1998), in their computer model, count selectional restrictions among important cues for disambiguating coordinate structures. They also refer to a statistics-based default in one of their algorithms (see below).

Now, I propose the following strategy:

Strategy (C)

In the pattern "X Co Y *no* Z<sub>1</sub>", whether or not X and Y correspond semantically, if "X *no* Z<sub>1</sub>" and "Y *no* Z<sub>1</sub>" are both normal co-occurrences and if they represent the same semantic relations between the nouns connected by *no*, then produce '[X Co Y] *no* Z<sub>1</sub>' as a default interpretation. If "Y *no* Z<sub>1</sub>" but not "X *no* Z<sub>1</sub>" is a normal co-occurrence or if the constraint of the same semantic relations is not satisfied, produce 'X Co [Y *no* Z<sub>1</sub>]' for the moment. In the pattern "W<sub>m</sub> *no* X Co Y" as well, if parallel conditions are satisfied, produce 'W<sub>m</sub> *no* [X Co Y]' as a default interpretation. If "W<sub>m</sub> *no* X" but not "W<sub>m</sub> *no* Y" is a normal co-occurrence or if the constraint of the same semantic relations is not satisfied, produce '[W<sub>m</sub> *no* X] Co Y' for the moment.

It may help to think of Strategy (C) as having two functions: regarding those patterns to which Strategy (A-1) has already been successfully applied, it checks whether X and Y can be treated as a whole unit to which modification applies; regarding those patterns to which neither Strategies (A) nor Strategy (B) has been successfully applied, it tentatively gives a promising interpretation, thereby satisfying the requirement of early attachment. In the latter case, it is a weak strategy in the sense that it comes into play only when the other strategies cannot handle the matter.

Now, it should be mentioned that the Japanese genitive marker *no* can express a wide variety of semantic relations between the nouns connected by it (Sasano et al. (2005: 131ff)), as the English preposition *of* or prenominal possessive (*'s*) can. For example, *Hanako no e* ('Hanako's picture') can mean a picture drawn by Hanako, one owned by Hanako, one in which Hanako is drawn, etc., depending on the context.

Let us first look at some examples to which Strategy (C) applies negatively.

(24) *Bahha ya Piasora no "Riberutango"* (*Asahi-sinbun*, 29/10/1998, Evening, 16)

Bach and Piazzolla GEN "Libertango"

'Bach and "Libertango" by Piazzolla'

In (24), although both *Bahha* and *Piasora* are in the same semantic category of composers, only the latter is considered to be related to "*Riberutango*" ('*Bahha ya [Piasora no "Riberutango"]*'), because Piazzolla but not Bach is known to be the author of "Libertango". The next two examples, in which X and Y do not quite correspond semantically, illustrate the same point:

(25) *Kenpoo to kokumin no koe* (A084558)  
 constitution and people GEN voice  
 'the constitution and the voice of the people'

(26) *Nenrei ya saigetū no nagare* (E: 124)  
 age and years GEN stream  
 'the (person's) age and the passage of years'

In (25), *koe* is only modified by *kokumin no* because *kenpoo no koe* 'the voice of the constitution' is an anomalous co-occurrence in Japanese. Similarly, in (26) *Nenrei no nagare* does not make sense in Japanese. The next example illustrates the constraint of the same semantic relations:

(27) *Kokuren- kinen- kitte to [Tōkyō- sibu no atukai]* (A053891)<sup>14</sup>  
 the UN commemoration stamp and Tokyo branch GEN treatment  
 'the UN commemorative stamps and the treatment by the Tokyo branch'

In (27), the first interpretation that comes to mind is the one indicated by bracketing, although *kokuren-kinen-kitte no atukai* 'the treatment of the UN commemorative stamps' is a perfectly normal co-occurrence. This may be explained as follows. When Strategy (C) is tried, being an inanimate object, *kitte* cannot be construed as the agent of *atukai*; it is most naturally construed as the patient. *Tōkyō-sibu*, on the other hand, can be either the agent or the patient of *atukai*, depending on the context. However, it is difficult to imagine a situation where both *kitte* and *Tōkyō-sibu* are the patients of *atukai* at the same time. Therefore, *Tōkyō-sibu* must be construed as the agent. Since the semantic relations between the nouns connected by *no* are not the same, Strategy (C) applies negatively.<sup>15</sup>

In contrast, several examples to which Strategy (C) applies positively are given below. In other words, their default interpretations, which are endorsed in the contexts, are the ones indicated by bracketing because the co-occurrences in question are normal, e.g., *Yōroppa no rentyū, kokoro no kiroku, gizyūto no kyōkashyō, geinōkai no kashyū*, etc., and because the constraint of the same semantic relations is satisfied:

(28) [*Yōroppa ya Amerika*] *no rentyū* (U: 191)  
 Europe and America GEN fellows  
 'European and American fellows'

(29) [*Kokoro ka tamasii*] *no kiroku* (U: 163)  
 heart or soul GEN record  
 'the record of one's heart or soul'

(30) *Gizyūto no [kaisetūsyō ya kyōkashyō]* (U: v)  
 technique GEN manual and textbook  
 'manuals and textbooks on techniques'

(31) *Geinōkai no [haiyū ya kashyū]* (E: 103)  
 the entertainment world GEN actor and singer

'actors and singers in the entertainment world'

- (32) [*Zyooahoo kensaku to hassoo*] *no sinsisutem* (N: 1)  
 information retrieval and conception GEN new system

'a new system of information retrieval and conception'

- (33) [*Buhin ya tokusei*] *no subete* (N: 46)  
 parts and qualities GEN all

'all of the parts and qualities'

- (34) *Syosai no [syodana ya taipuraitaa]* (S: 129)  
 study GEN bookshelf and typewriter

'the bookshelf and typewriter in the study'

In (28)~(31) X and Y correspond semantically but in (32)~(34) they do not.

At this point I should state beforehand the relationships between the strategies (Table 1).

		A-2	B	C
A-1	ordering	=	>	>
	ranking	-	>	>
A-2	ordering		>	>
	ranking		-	-
B	ordering			=
	ranking			>

Table 1

#### Relationships between Strategies in terms of Ordering and Ranking

By *ordering* I mean a logical order in which the strategies are applied: in Table 1, ">" means that a successful or unsuccessful application of a strategy on the left side is presupposed for a subsequent application of another strategy on the upper side; "=" means that the strategies in question may be tried in parallel or at least that logically it is difficult to say which is applied first. Most of the orderings follow from the very formulations of the strategies: the application of Strategy (B), for example, occurs only after a full appearance of the corresponding heads of the coordinated noun phrases, which means that Strategy (A-1) and Strategy (A-2) have already been applied with or without success. By *ranking* I mean the strength of one strategy in relation to another, on which the former wins the competition with the latter, either while the two are being applied in parallel or after the latter has already been applied. In Table 1, ">" means that a strategy on the left side is stronger than another strategy on the upper side (I will substantiate these claims in Subsection 4.3.); "-" means that ranking is irrelevant between the strategies in question or at least that should there be some ranking between them, no appropriate example has been discovered in the corpora or can be invented which could



substantiate it.

## 4.2 Details

In this subsection, additional information about the strategies is presented.

(A) The strategies based on corresponding semantic categories

First, let us examine a few longer examples.

(35) *Itabasiku- Narimasu to tonari no dooku- Kamiakatuka* (J137717)

Itabasi ward Narimasu and neighbor GEN the same ward Kamiakatuka

'Narimasu in Itabasi Ward and neighboring Kamiakatuka in the same ward'

(36) *Tyuugakusei ya kookoo- svogakunen no seito* (E: 206)

junior high school student and senior high school first year GEN student

'junior high school students and first-year students of senior high school'

In (35), Strategy (A-1) is called on to see at first whether or not X corresponds to Y (*tonari*). Because the answer is in the negative, it then examines the semantic category of Z<sub>1</sub> (*dooku-Kamiakatuka*), with the result that Z<sub>1</sub> turns out to correspond to X semantically. Analytically speaking, the first constituent of X (*Itabasiku*) and that of Z<sub>1</sub> (*dooku*) correspond to each other, and the second constituent of X (*Narimasu*) and that of Z<sub>1</sub> (*Kamiakatuka*), being the names of the streets, correspond to each other. In (36), X corresponds not to Y but to "Y no Z<sub>1</sub>" as a whole. The point is that the structural ambiguities should be resolved in such a way as to manifest the correspondence of semantic categories.

A question may duly be asked at this point as to what it means exactly to say that a noun corresponds to another semantically. As Miller and Charles (1991: 2) state, there is a gradation in semantic similarity between any two nouns. Instead of giving a definitive answer to the question, I will show different types of inherent semantic similarity to such a degree that a pair of nouns can be judged to be the correspondents of coordinate noun phrase structures, quite independently of context. Note that the aim is not to give a complete and systematic classification; it is rather to show some further examples of a great semantic similarity, thereby giving some idea of how I proceed to apply Strategies (A) to the data in the evaluation process, which will be mentioned in Section 6. Gray zone cases, after all, may be taken care of by Strategy (C) above.

(a) Fellowship relations

Near synonyms:

(37) *Mukasi no sirvoo ya deeta* (N: 48)

old times GEN materials and data

'old materials and data'

Co-hyponyms (common nouns):

(38) *Piano ya baiorin no ensoo* (U: 10)

piano and violin GEN playing

'the playing of the piano and violin'

Co-hyponyms (proper nouns):

(39) *Oo* to *Nagasima* no *sebangoo* (S: 138)

Oo and Nagasima GEN player's uniform number

'Oo's and Nagasima's uniform number'

Frame-based fellow terms:

(40) *Kozutumi* no *uketuke*- *madoguti* to *kitte* no *madoguti* (A016463)

parcel GEN reception window and stamp GEN window

'the parcel post window and the stamp window'

The same first constituents:

(41) *Zyuumin*- *ziti* no *kakudai* to *zyuumin*- *hukusi* no *koozyoo* (C143714)

resident self-government GEN expansion and resident welfare GEN improvement

'the expansion of residential self-government and the promotion of residential welfare'

It is easy to see that both synonyms and co-hyponyms are a pair of nouns with a sufficient semantic similarity to be targeted by Strategies (A). By frame-based fellow terms I mean that a pair of nouns which do not inherently share so many semantic features can be regarded as fellows in terms of a given frame in the sense of Fillmore (1982: 111): any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits. For example, although a parcel and a stamp, viewed in themselves, are not physically very similar objects, they are considered as fellows in the sense that they are both treated in a post office. As for the same first constituents, the idea is that, if a pair of compound nouns have the same morpheme in their first position, it gives them an appearance of being in the same semantic category, although in Japanese the semantic category of a compound noun as a whole generally coincides with that of its last constituent rather than its first constituent.

(b) Opposing relations

Contrasting terms:

(42) *Sakumotu* no *ryoo* to *situ* (A068242)

crops GEN quantity and quality

'the quantity and quality of the crops'

Inherent opposites:

(43) *Ama* to *puro* no *ryoohoo* (S: 139)

amateur and professional GEN both

'both amateurs and professionals'

Frame-based opposites:

(44) *Umi* to *sora* no *pareedo* (K065775)<sup>16</sup>

sea and sky GEN parade  
 'parades in the sea and in the sky'

Opposites in a set phrase:

- (45) *Syumi* to *zitueki* no *keiyaku-siiku* (C107596)  
 hobby and profit GEN contract breeding  
 'the contract breeding at once for pleasure and for profit'

I follow Cruse (1986) in using the term *opposites* as a cover term comprising antonyms, antipodals, converses, etc. Opposites are very easily recognized as semantically corresponding elements because they are a special case of co-hyponyms and typically have all but one semantic features in common. The distinction between contrasting terms and opposites is not intended to be definite. As Cruse (1986: 197ff.) states, opposites are not a well-defined class, and there are peripheral instances as well as prototypical ones. As for opposites in a set phrase, the idea is that if a pair of nouns with somewhat contrastive meanings are used as part of a cliché, they can easily be recognized as semantic correspondents elsewhere; for example, many Japanese dictionaries list the phrase *syumi to zitueki o kaneru* 'combine fun and practicality'.

(B) The strategy based on binding

First of all, in order to deal with more complex structures, the strategy should be stated in general terms as follows:

In the pattern " $W_1$  no  $W_2$  no..... $W_m$  no X to Y no  $Z_1$  no  $Z_2$  no..... $Z_n$ ", if  $Z_n$  represents a concept that presupposes plural entities besides itself, then create a combinatory interpretation such that both X and  $Z_{n-1}$  are the heads of the coordinated noun phrases and  $Z_n$  is the head of the whole pattern.

Let us illustrate this with an example adapted from (14), with a noun added at the end:

- (46) Honkon no syooboosi to Tai no keisatukan no kutigenka  
 Hong Kong GEN fireman and Thailand GEN police officer GEN quarrel  
 'a quarrel between a fireman from Hong Kong and a police officer from Thailand'

In (46), the key noun is  $Z_2$  (*kutigenka*). Accordingly, (46) is interpreted as '<[*Honkon no syooboosi*] to [*Tai no keisatukan*] > no *kutigenka*', *syooboosi* and *keisatukan* being the corresponding heads.

Second, this strategy is applicable only in cases where Co is none other than *to*, so that in the following example combinatory interpretation of coordinated nouns or noun phrases does not occur in spite of the presence of two underlined key nouns:

- (47) *Sizen-kagaku* ya *gizyutu* no kankei no *hitotati* no aida de (U: 185)  
 natural science and technique GEN relation GEN people GEN interval LOC  
 'among people engaged in natural science and technology'

Third, when a pair of nouns are recognized as the corresponding elements of coordination by

applying this strategy, their correspondence sometimes exists mainly on the strength of the semantic feature of the adjacent key noun; unlike in the case of Strategies (A), they do not have to be inherently similar except on a fairly abstract level:

- (48) *Sinbunsi to boku no aida* (Dan, I., *Paipu no Kemuri*: 22)  
 newspaper and I GEN interval  
 'between the newspaper and me'

In (48), *sinbunsi* and *boku* have in common only that they are both physical objects. In fact the similarity could be even less. For example, the noun *kankei* 'relation' only requires that the corresponding nouns should have a possibility of being related somehow, so that almost any pair of nouns will do, e.g., *ringo to gengogaku no kankei* 'the relation between an apple and linguistics'. The correspondence established in this non-taxonomic manner may be called abstract correspondence as contrasted with semantic correspondence.

(C) The strategy based on co-occurrence preferences

First of all, I would like to elaborate on the motivation to postulate a default interpretation specified in Strategy (C) (in order to do this, I have to extend my argument for a while to a set of more general structures formulated as (3)). It has rightly been pointed out that coordinate noun phrase structures of the "old men and women" type are ambiguous, at least out of context. But it is another question whether there is no strong tendency for them to be interpreted in one way rather than the other in context. Regarding "old men and women" Quirk et al. (1985: 960) state that the usual interpretation is 'old [men and women]' but that there is also the possibility of interpreting it as '[old men] and women'. Huddleston and Pullum (2002: 1286) also state that, other things being equal, the structure with the modifier applying to the coordination is in general more likely.<sup>17</sup> These statements may be enough for the English usage but a corresponding statement is far from being satisfactory in the case of Japanese. In fact, Yamanaka (2000) finds that when "W'X" and "W'Y" are normal co-occurrences and in addition X and Y are in the same semantic category (thus covered by Strategy (A-1)), as in (49) below, examples of the "W' X Co Y" type are almost always (50 times) interpreted as 'W' [X Co Y]' in context, with only one clear exception (*nihiki no inu to usagi* (gl. 'two head GEN dog and rabbit') 'two dogs and a rabbit' (E: 80)).

- (49) *Miryokutekina onna ya otoko* (S: 36)  
 attractive woman and man  
 '[attractive women] and men' or 'attractive [women and men]'

Further, Yamanaka (2000) also finds that a great majority of the examples of the "W' X Co Y" type (117 out of 131), where X and Y are not in the same semantic category, are interpreted as 'W' [X Co Y]' in actual context, as long as "W'X" and "W'Y" are normal co-occurrences. Here I may appear to beg the question by bringing up contextual matters in discussing the strategies applied *out of context*, but it is not so. The point is simply this: by postulating a default

interpretation, no extra mechanism is needed to explain the observed bias in interpretation in contexts, which are largely more or less neutral ones forcing neither the one nor the other interpretation.

Next, regarding the aforementioned various types of anomaly engendered by non-satisfaction of co-occurrence preferences, I would like to draw a distinction between collocational, semantic, and pragmatic one.<sup>18</sup> This distinction is mainly intended to serve instructive purposes, and not to exclude other systems of classification or to deny the possibility that various degrees of anomaly may in fact form a continuum, as suggested by Cruse (2004: 219).

Collocational anomaly does not consist in the clash between the propositional meanings of the co-occurring units but results from non-satisfaction of a somewhat arbitrary stylistic presupposition:

- (50) *Basabasa no atama to hige* (E: 179)  
 unkempt GEN head and beard  
 'unkempt hair and a beard'

*Basabasa* is a mimetic word normally used to describe hair but not a beard; *booboo* 'unkempt, untrimmed' can be used for both hair and a beard. Even if a person's beard were of the same shape as his/her hair appropriately described as *basabasa*, we would still hesitate to describe the former as *basabasa*.<sup>19</sup> Hence, (50) is interpreted as '[*basabasa no atama*] to *hige*'

Semantic anomaly results from the clash between the propositional meanings of the co-occurring units:

- (51) *Sinkan no syoten to huruhon'ya* (S: 224)  
 new publications GEN bookstore and used-book store  
 'bookstores of new publications and used-book stores'

*Sinkan no huruhon'ya* 'used-book stores of new publications' would involve a contradiction in normal circumstances, so that (51) is interpreted as '[*sinkan no syoten*] to *huruhon'ya*'.

Pragmatic anomaly has to do with non-satisfaction of expectations regarding prototypical situations in the extralinguistic world:

- (52) *Zisatusita Hara-Tamiki-si to boku* (E: 37)  
 committed suicide Hara Tamiki Mr. and I (male)  
 'Mr. Hara Tamiki, who committed suicide and I'

There is nothing wrong with *zisatusita boku* 'I, who committed suicide' semantically, but pragmatically it requires a rather unusual context for this co-occurrence to be used appropriately, as when somebody refers to himself in his suicide note. Hence, in most cases (52) is interpreted as '[*zisatusita Hara-Tamiki-si*] to *boku*'.

Now, Strategy (C) should be stated more formally and generally as follows because its formulation in Subsection 4.1 is informal and suited to simple cases such as (24)~(34).

Strategy (C) (general version)

When candidate corresponding nominals are discovered, examine the co-occurrence preferences of each of the nominals on the one hand and the elements with which that nominal is in syntactically possible modifying relation on the other. If the co-occurrence proves to be anomalous on semantic, pragmatic, or collocational grounds, or if the constraint of the same semantic relations is not satisfied, exclude those interpretations which imply either. If the co-occurrence is normal and the constraint of the same semantic relations is satisfied, produce as a default interpretation the one in which the element in question is in modifying relation with both of the nominals.

For sometimes "X Co Y" modifies not  $Z_1$  but " $Z_1$  no  $Z_2$ " as a whole or a noun phrase farther to the right: sometimes the correspondent of X is not Y but " $Y$  no  $Z_1$ " as a whole or a noun phrase farther to the right. The example below illustrates the latter case:

- (53) *Sensei ya annai no boosama no setumei* (E: 206)  
 teacher and guidance GEN bonze GEN explanation  
 'explanations by a teacher and a ushering bonze'

In (53), the correspondent of *sensei* is " $Y$  no  $Z_1$ " (*annai no boosama*) as a whole with *boosama* as its head. Accordingly, Strategy (C) examines the appropriateness of the co-occurrences " $X$  no  $Z_2$ " (*sensei no setumei*) and " $Z_1$  no  $Z_2$ " (*boosama no setumei*). Because both co-occurrences are normal and the constraint in question is satisfied, (53) is usually interpreted as '[*sensei ya [annai no boosama]*] no *setumei*'.

Finally, as in the case of Strategy (B), when a pair of nouns are recognized as the correspondents of coordination by applying Strategy (C), their correspondence may exist on the strength not so much of their inherent semantic similarity as of the semantic feature of the adjacent noun; it may involve abstract correspondence:

- (54) *Hakumai to enbun no torisugi* (E: 182)  
 polished rice and salinity GEN too much intake  
 'too much intake of polished rice and salinity'

In (54), the semantic similarity between the corresponding nouns is a rather vague one, i.e., 'something to be ingested', and it is exactly the reflection of the meaning of the following noun *torisugi*.

### 4.3. Interaction: analyses of complex examples

So far, in illustrating the strategies, we have restricted ourselves to relatively simple structures. Let us now look at some of the more complex structures to see how the strategies are made to work consecutively, and consider their relationships. In so doing, I will also be proposing some hypotheses on incremental processing of linguistic input. The complex structures examined here fall into two categories, depending on whether or not they involve error recovery.

(i) Examples which do not involve error recovery

Examples in this category are not really complex in the true sense but only lengthy; they are interpreted straightforwardly by applying the strategies consecutively. Let us study the following:

- (55) *Mae no tatemono no ikkai no baa no aka to ki no neon* (S: 38)  
front GEN building GEN first floor GEN bar GEN red and yellow GEN neon  
'the red and yellow neon signs of the bar on the first floor of the building in front'

In (55), the first problem is to determine which noun or noun phrase *mae no* is going to modify. Since *mae no tatemono* 'the building in front' is a normal co-occurrence, in accordance with the strategy proposed in Yamanaka (1988) (see (7)-(i) in Subsection 4.1), the interpretation '[*mae no tatemono*] *no ikkai*' is arrived at up to that point. The same strategy is applied repeatedly to give the interpretation '[ [*mae no tatemono*] *no ikkai*] *no baa*'. However, at the next stage the application of this strategy is suspended, because *baa no aka* is not a fully natural collocation or at least does not make clear exactly in what sense it is to be interpreted. The first noun after the coordinator *to* is *ki*, which is in the same semantic category as *aka* and which is readily taken as its correspondent, according to Strategy (A-1). Finally, when the next noun *neon* is encountered, co-occurrence preferences are examined as required by Strategy (C) (whether *aka no neon* and *ki no neon* are normal co-occurrences) and by Strategy (7)-(iii) in Subsection 4.1 (whether *baa no neon* is all right). As the co-occurrences are normal, the end result is the interpretation '[[ [*mae no tatemono*] *no ikkai*] *no baa*] *no* [[*aka to ki*] *no neon*]', which is in fact the one native speakers make of (55).<sup>20</sup>

(ii) Examples which involve error recovery

In the case of reading as opposed to listening, sentence comprehension process, especially error recovery depends very much on the quantity of linguistic information which we can get at a glance and process, that is, the number of letters/characters or words.<sup>21</sup> As this in turn depends on many factors, such as the speed of reading, the perceived size of letters, etc. and as I know of no suitable literature on which to base my hypotheses, I make the usual assumption that linguistic information is presented to us word by word. Let us examine minutely the following:

- (56) *Sihoo-seido no zyunka to kensatukan no sekinin no meikakuka* (A083064)  
judicial system GEN purification and prosecutor GEN responsibility GEN demarcation  
'the purification of the judicial system and the demarcation of the prosecutor's responsibility'
- (57) *Sihoo-seido no zyunka to kensatukan.....*

At this point, since neither *sihoo-seido* nor *zyunka* is in the same semantic category as *kensatukan*, Strategies (A) are not applicable. Strategy (C) is then tried but without success, because *sihoo-seido no kensatukan* is a slightly anomalous collocation, and in any event does not represent the same semantic relation as *sihoo-seido no zyunka*. This leaves us with the

interpretation '[*sihoo-seido no zyunka*] to *kensatukan*' ('[the purification of the judicial system] and the prosecutor') up to this point, which would be the final one without further input.

(58) *Sihoo-seido no zyunka to kensatukan no sekinin*.....

When the next noun *sekinin* is encountered, Strategies (A) are tried again but they fail again for the same reason as before. Then Strategy (C) is tried again in order to see if it is possible to interpret the phrase as '[[*sihoo-seido no zyunka*] to *kensatukan*] no *sekinin*' or '*sihoo-seido no [zyunka to [kensatukan no sekinin]]*'. Neither interpretation is hardly possible: in the former case, *zyunka no sekinin* 'the responsibility for purification' and *kensatukan no sekinin* 'the prosecutor's responsibility' are not parallel, thus violating the constraint of the same semantic relations imposed on Strategy (C); in the latter *sihoo-seido no zyunka* and *sihoo-seido no sekinin* are not parallel semantically, and in addition '*sihoo-seido no [kensatukan no sekinin]*' is not a fully natural collocation.<sup>22</sup> We cannot but interpret (58) as '[*sihoo-seido no zyunka*] to [*kensatukan no sekinin*]' ('the purification of the judicial system and the prosecutor's responsibility'). To return to (56), the next noun *meikakuka* is in the same semantic category as *zyunka*, which allows Strategy (A-1) to be applied to give the final interpretation '[*sihoo-seido no zyunka*] to [[*kensatukan no sekinin*] no *meikakuka*]'. With this example, we could say that it involves error recovery in the sense that the noun corresponding to *zyunka* shifts from *kensatukan* via *sekinin* to *meikakuka* as new input comes in. But as error recovery occurs immediately, it is very probable that we are hardly conscious of having made an error, unlike in the case of those garden-path sentences in which there are a few intervening words between the error and its recovery.

Granting that (56) involves error recovery, it suggests that Strategy (A-1) takes precedence over Strategy (C): the former candidate for the counterpart of *zyunka*, selected negatively by Strategy (C), i.e., *sekinin*, is beaten by a newcomer recommended solely by Strategy (A-1), i.e., *meikakuka*, with no further check being needed.

In this connection, I would like to make a few claims regarding the ranking, or relative strength of other pairs of strategies. First, Strategy (A-1) also seems to take precedence over Strategy (B). Let us look at a likely example adapted from (79) below:

(59) *Tikara no sa to gizyutu no sa*  
 power GEN discrepancy and skill GEN discrepancy  
 'discrepancy in power and discrepancy in skills'

In this example, Strategy (B) could be triggered by the key noun *sa* in its second occurrence, leading to the nonsensical interpretation '<[*tikara no sa*] to *gizyutu*> no *sa*' 'discrepancy between discrepancy in power (on the one hand) and skills (on the other)'. But priority is given to Strategy (A-1), which produces the correct interpretation '[*tikara no sa*] to [*gizyutu no sa*]'.<sup>23</sup>

As for Strategy (B) versus Strategy (C), the former seems to be stronger. Let us look at a likely example adapted from (74) below:



- (60) *Otoko to onna no yuuzyoo*  
 man and woman GEN friendship

Since *otoko* and *onna* belong to the same semantic category, Strategy (A-1) has successfully been applied by the time *yuuzyoo* appears. Then Strategy (C) is called on to check whether *otoko no yuuzyoo* and *onna no yuuzyoo* are normal co-occurrences, with the result that (60) is interpreted as [*otoko to onna*] *no yuuzyoo* ('friendship among men and friendship among women'). But there is another, more plausible, interpretation of (60), that is, '<*otoko to onna*> *no yuuzyoo*' ('friendship between a man and a woman'). This is possible because *yuuzyoo*, being a key noun, triggers the application of Strategy (B).

Let us take another example of possible error recovery in order to argue another point:

- (62) *Miura no zinkeikan to isi no tuyosa* (E: 168)  
 Miura GEN view of life and will GEN strength  
 'Miura's view of life and the strength of his will'

- (63) *Miura no zinkeikan to isi*.....

At this point, Strategy (A-1) is not applicable because we are not sure that *zinkeikan* and *isi* belong to the same semantic category. Then the applicability of Strategy (C) is examined: *Miura no isi* 'Miura's will' is a perfectly normal co-occurrence; *Miura no zinkeikan* and *Miura no isi* represent almost the same semantic relations in the sense that both are paraphrasable as "something which is part of Miura's personality". Accordingly, (63) can be interpreted as '*Miura no [zinkeikan to isi]*'.

- (64) *Miura no zinkeikan to isi no tuyosa*

The next noun is *tuyosa*, which is not in the same semantic category as *Miura* or *zinkeikan*, thus rendering Strategies (A) inapplicable. Then again Strategy (C) is tried. But this time (64) cannot be interpreted as '*[Miura no [zinkeikan to isi]] no tuyosa*': although *isi no tuyosa* 'the strength of will' is a perfectly normal co-occurrence, *zinkeikan no tuyosa* 'the strength of one's view of life' is not. So we must look for another way of interpreting (64). *Zinkeikan* on the one hand and *isi no tuyosa* on the other can be good correspondents, both belonging to Miura's personality. Does it amount to saying that we must interpret (64) as '*Miura no [zinkeikan to [isi no tuyosa]]*', thereby canceling the modification of *isi* by *Miura no* formed at an earlier stage, that is, treating it as an error? Not necessarily. Our mental representation built in interpreting (64) can be written, in my view, graphically as follows:

- (65) '*Miura no /zinkeikan to [isi] no tuyosa*'<sup>24</sup>

All (65) means is this: both *zinkeikan* and *isi* belong to Miura; the counterpart of *zinkeikan* is not *isi* but *isi no tuyosa* (only in this sense there is error recovery). That seems to be exactly what we understand in processing (64), for the will in question cannot possibly belong to anyone other than Miura. I do admit that such a representation as (65) looks grotesque or even absurd in the eyes of the majority of linguists because it does not conform to conventional

formalism. But it is no more than an assumption yet to be verified to think that our mental representations in language understanding must be formed, in each and every case, in the manner corresponding to conventional syntactic representations which linguists construct with a pencil on paper and which are suitable for computational processing, and I would like to question that very assumption. This last point may perhaps be better illustrated with a rather lengthy but more revealing example:

(66) *Watasi no yomitakatta Zyoan-Didian ya Zyoosi-Purinputon no hon nado* (S: 236)

I GEN wanted to read Joan Didian and George Primpton GEN book etc.

'books by Joan Didian and by George Primpton, which I wanted to read, etc.'

(67) *Watasi no yomitakatta Zyoan-Didian ya Zyoosi-Purinputon.....*

'Joan Didian and George Primpton, who I wanted to read'

At this point Strategy (A-1) applies because *Zyoan-Didian* and *Zyoosi-Purinputon* (henceforth *Z-D* and *Z-P*, respectively) are both personal names. Then Strategy (C) is tried. Both *watasi no yomitakatta Z-D* and *watasi no yomitakatta Z-P* are normal co-occurrences as long as we interpret *Z-D* and *Z-P* metonymically as *Z-D*'s and *Z-P*'s writings, and this metonymical extension of meaning is common enough in Japanese as well as in English, e.g., *read Shakespeare*. The next noun is *hon* and again the application of Strategy (C) gives a positive result: both *Z-D no hon* 'books by Joan Didian' and *Z-P no hon* 'books by George Primpton' are perfectly normal co-occurrences. In this process *Z-D* and *Z-P* cease to have metonymical meanings, and in this restricted sense error recovery can be said to have occurred. Our interpretation of (66) so far could plausibly be shown as :

(68) '[*Watasi no yomitakatta* {[*Z-D ya Z-P*] *no hon*}]'.<sup>25</sup>

Now a problem arises: strictly speaking, is it not more appropriate to interpret (66) as '*watasi no yomitakatta* [[*Z-D ya Z-P*] *no hon*]', because the object of reading is not a person but a book? Was it not an error, accordingly, to think at first that *watasi no yomitakatta* modifies *Z-D* and *Z-P*? This kind of argument seems to be *ex post facto*. That is, it is only in retrospect that we find that *Z-D* and *Z-P* have literal meanings; in real time understanding, there is nothing to prevent us from interpreting them metonymically as their writings. Perhaps more importantly, there is little, if any, difference in meaning whether we interpret (66) as '[*watasi no yomitakatta* {[*Z-D ya Z-P*] *no hon*}]' or as '*watasi no yomitakatta* [[*Z-D ya Z-P*] *no hon*]' : we understand in either case that there are certain books *I* wanted to read and that they are written by *Z-D* and *Z-P*. The psychological implausibility of assuming error recovery or suspension of interpretation as to modification in this example is further strengthened by the fact that there is a long distance between the end of the modifying clause and the last noun *hon*, which is the modified one from the logical point of view and which putatively triggers error recovery.<sup>26</sup>

This way of conceiving modification flexibly enables the system to reduce error recovery to a

minimum, which in fact seems to be the case.<sup>27</sup> If I am right on this point, it will suggest that our mental representations in language processing do not always coincide with syntactic representations in the traditional analysis.<sup>28</sup>

In this connection, it will lend further support to the above claim to consider a few examples of what Quirk et al. (1985: 972) call grammatically 'ill-assorted' coordination:

(69) We have washed, dried, and put the dishes away.

(70) By giving the police a pay rise, the Minister hopes to strengthen and make the force more efficient.

Quirk et al. claim that *away* in (69) and *more efficient* in (70) are structurally outside the coordinate construction but that they semantically apply only to the last conjoin. Quirk et al. state also that there is no difficulty in understanding such sentences and that it is not clear how far they are to be regarded as being ungrammatical in a descriptive sense, or as being stigmatized merely because of prescriptive tradition. The problem is that, although they are understandable semi-grammatical sentences, it is difficult to assign them correct syntactic representations even in the retrospective analysis.

## 5. Problematic examples

In the body of examples I examined, there are some which are problematic in the sense that their actual interpretations differ from those predicted by the strategies. They fall into two categories, depending on whether the factors responsible for the unpredicted interpretation lie within the ambiguous region or outside it, that is, in the wider context.<sup>29</sup>

I will mainly discuss those cases which exhibit a pattern, rather than list all the problematic examples I encountered, some of which are trivial or idiosyncratic.

(i) The factors within the ambiguous region

(71) *okusan to okusan no ryoosin no yonin-gurasi* (A051487)

wife and wife GEN parents GEN four persons living

'a family of four — himself, his wife, and his wife's parents — living together'

(72) *zibun to zibun no koibito no zyoozi* (E: 143)

self and self GEN lover GEN amour

'one's love affair with one's lover'

In this type of examples, where the very same nouns are connected by a coordinator, Strategy (A-1) is immediately called on to label them as semantically corresponding nouns. The next step is for Strategy (C) to check whether "X *no* Z<sub>1</sub>" and "Y *no* Z<sub>1</sub>" are normal co-occurrences. In these examples the co-occurrences in question (*okusan no ryoosin* '(his) wife's parents'; *zibun no koibito* 'one's lover') are quite normal. According to Strategy (C), the default interpretations of (71) and (72), then, are '[[*okusan to okusan*] *no ryoosin*] *no yonin-gurasi*' and '[[*zibun to zibun*] *no koibito*] *no zyoozi*', respectively, which differ from their actual interpretations, i.e., '<*okusan*

to [okusan no ryoosin]> no yonin-gurasi' and '<zibun to [zibun no koibito]> no zyoozi'. This problem may be taken care of by extending the concept of anomalous co-occurrences under Strategy (C), thereby depriving (71) and (72) of the status of real counterexamples to Strategy (C). That is, to take (71) for example, although *okusan no ryoosin* in itself is a normal co-occurrence, the co-occurrence mediated by *no* of *okusan to okusan* with *ryoosin* is pragmatically anomalous: if both of the *okusans* refer to the same person, one of them is redundant and should have been spared; if they refer to different persons, some measures should have been taken to make the distinction clear, e.g., *kono okusan to ano okusan no ryoosin* 'the parents of *this* wife and *that* wife'. In contrast to (71) and (72), there are many examples of "X to Y no Z<sub>1</sub>" (,where X and Y have the same intensional meaning, albeit a different extension) which are interpreted as '<X to Y> no Z<sub>1</sub>' and not 'X to [Y no Z<sub>1</sub>]' :

- (73) *Me to me no kankaku* (K054439)  
 eye and eye GEN distance  
 'the distance between the eyes'
- (74) *Otoko to otoko no yuuzyoo* (*Asahi-sinbun*, 11/11/1998, Morning, 22)  
 man and man GEN friendship  
 'friendship among men'
- (75) *Otoko to otoko no aida ni* (S: 37)  
 man and man GEN interval LOC  
 'between one man and another'
- (76) *Kaado to kaado no aida ni* (U: 58)  
 card and card GEN interval LOC  
 'between cards'
- (77) *Go to go no aida* (U: 190)  
 word and word GEN interval  
 'between words'

However, these unproblematic examples are, as far as I have been able to ascertain, restricted to a subtype in which Z<sub>1</sub> represents a concept that presupposes plural entities besides itself (see Strategy (B) in Subsection 4.1). In sum, there is regularity in the interpretation of examples of the "X to X no Z<sub>1</sub>" type: they are interpreted either legitimately as '<X to X> no Z<sub>1</sub>' in accordance with Strategy (B) or as 'X to [X no Z<sub>1</sub>]' by recourse to the extended concept of pragmatic anomaly under Strategy (C).

- (78) *Sei to suu no kubetu* (U: 131)  
 Gender and number GEN distinction  
 'distinctions as to gender and number'
- (79) *Tikara to gizyutu no sa* (C155458)  
 power and skill GEN discrepancy

'discrepancy in power and in skills'

In (78), *kubetu* is a special noun which invites the application of Strategy (B), but in context (78) is not interpreted as '<*sei to suu*> *no kubetu*' ('the distinction between gender and number') but as '[*sei to suu*] *no kubetu*'. This unexpected interpretation may be due to knowledge accessible through (78) itself: there is a certain distinction as to gender (masculine, feminine, etc.) and also a certain distinction as to number (singular, plural, etc.), and there are cases in which both distinctions are relevant, as in the case of the category of nouns in natural language. In fact, it might be only on the strength of the background knowledge that certain categories have distinctions both as to gender and as to number that the intended interpretation of (78) can be arrived at. Linguists may have no difficulty in arriving at it without any wider context, but some people may be at a loss how to interpret it out of context and they may perhaps end up in interpreting (78) as '<*sei to suu*> *no kubetu*' simply by applying Strategy (B) to it. Likewise, in (79) *tikara* and *gizyutu* conjure up in our mind a competitive situation in which discrepancy both in power and in skills is relevant. In sum, these apparent counterexamples should not be taken to undermine Strategy (B) itself. Let us look at reverse cases, in which apparently unexpected application of Strategy (B) is called for:

(80) *Teisei to owabi no bunsyo* (A076715)

correction and apology GEN note

'a note (at once) for correction and for an apology'

(81) *Kami to enpitu no huro-memo* (N: 195)

paper and pencil GEN bath memo

'a memo written in the bathroom in pencil on paper'

Neither *bunsyo* nor *memo* is a good candidate for a special noun which invites the application of Strategy (B), yet (80) and (81) are interpreted in context as '<*teisei to owabi*> *no bunsyo*' and '<*kami to enpitu*> *no huro-memo*' rather than '[*teisei to owabi*] *no bunsyo*' and '[*kami to enpitu*] *no huro-memo*', respectively and they are likely to be so interpreted even out of context. The key knowledge is this: in (80) correction and an apology presuppose a mistake, and if you make a correction and an apology for a mistake, you usually make them both at the same time. In other words, *teisei* and *owabi* jointly conjure up a frame in which a correction and an apology are made at the same time in some medium, and *bunsyo* denotes precisely such a medium. Likewise, in (81) pencil and paper are complementary in producing writing such as a memo.

Finally let us look at a few examples involving Strategy (C).

(82) *Sigoto no [kankyoo ya naiyoo]* (N: 152)

work GEN environment and content

'the environment for and the content of the work'

(83) *Umesakisi no [bungaku to hitogara]* (E: 116)

Mr. Umesaki GEN literature and personal character

'Mr. Umesaki's literature and personal character'

Strategy (A-1) cannot apply to these examples, because X and Y are not in the same semantic category. The positive application of Strategy (C) should also be blocked, because the constraint of the same semantic relations is not satisfied, although the co-occurrence preferences are respected. That is, in the case of (82) environment is one of the external conditions of the work, while content is its inherent quality and in the case of (83) literature is what Mr. Umesaki produces, while personal character is what he has as part of himself. However, the probable interpretations of (82) and (83), which are endorsed in the contexts, are that the modifier relates to both X and Y, as indicated by bracketing, so that it looks as though Strategy (C) has applied positively in spite of the breach of the constraint in question.

These apparent counterexamples can be taken care of by recourse to the concept of conceptual dependence inherent in certain nouns (Sasano et al. (2005: 130)). That is, *naiyoo* in (82) and *hitogara* in (83) have in common the feature that conceptually they do not usually stand alone but belong to something/body, i.e., the content of *something* and *somebody's* personal character. This feature makes the interpretations in question almost inevitable. If we interpreted (82) and (83) instead as '[*sigoto no kankyoo*] *ya naiyoo*' and '[*Umesakisi no bungaku*] *to hitogara*', respectively, they would hardly make good sense because it would then not be clear what content or whose personal character is being talked about, or to put it differently, there would be a striking imbalance in specificity between the conjuncts.

To sum up, the "problematic" examples discussed above may not be problematic in the true sense, because their irregularities can be accounted for by independent semantic or pragmatic factors. At least, they should not be taken to compromise the strategies themselves. In the evaluation of the strategies addressed in Section 6, however, they are subsumed under the category of problematic examples.

(ii) The factors in the wider context

(84) *Daigakuin no gakusei ya sutahhu* (N: 171)  
graduate school GEN student and staff

With regard to (84), whether or not *gakusei* and *sutahhu* are semantically similar enough, Strategy (C) is called on to examine co-occurrence preferences, etc. As *daigakuin no gakusei* 'graduate students' and *daigakuin no sutahhu* 'the staff of graduate school' are normal co-occurrences, and the semantic relations are the same, Strategy (C) allows us to interpret (84) as '*daigakuin no [gakusei ya sutahhu]*' ('[students and the staff] of graduate school'). However, there is something in the context which is incongruous with this interpretation:

(85) *Amerika no daigaku ni wa, koohii-meekaa no oite aru "tamariba" ga ari, daigakuin no gakusei ya sutahhu ga tatiyotte giron site iru.*

'In American universities, there are some rendezvous where one can find coffee-makers, and graduate students and the staff drop in there to have a discussion.'

Here the topic is American universities in general; restricting the staff to that of graduate schools as entailed by the interpretation in question is not encouraged, let alone necessitated, by the context. Instead, (84) in this context is preferably interpreted as '[*daigakuin no gakusei*] *ya sutahuu*' 'graduate students and the staff'.

(86) *Natume-Sooseki to desitati no atumari* (N: 168)

Natume Sooseki and disciples GEN gathering

In this example, *atumari* (lit. 'gathering') is a noun representing a concept which presupposes plural entities. Hence, by applying Strategy (B), (86) is most preferably interpreted as '<*Natume-Sooseki to desitati*> *no atumari*' ('the gathering of Natume Sooseki and his disciples').

Let us look at the immediately preceding context:

(87) *Keizai-gakusya-Keinzu ga, wakai toki ni Buruumuzuberii-guruupu to iu kootooteki-tisikizin no atumari ni sanku site ita koto, mata, koonen, kare no mawari ni Keinzu-saakasu to iu gakusya no atumari ga tukurareta koto wa yuumei de aru. Natume-Sooseki to desitati no atumari mo onazi da.*

'It is well known that Keynes, an economist, joined a circle of unworldly intellectuals called Bloomsbury Group when he was young and that afterward was formed around him a circle of scholars called Keynes's Circus. Under similar circumstances were Natume Sooseki and the circle of his disciples.'

Here Keynes and the circles which he was in are described as related but distinct entities. A parallel description in the interpretation of (86) is called for by the predicate *onazi da* 'be similar'. Hence, the interpretation of (86) in this context should be '*Natume-Sooseki to [desitati no atumari]*' 'Natume Sooseki and the circle of his disciples'.

To be precise, in order to arrive at the suitable interpretation we do not have to wait until the crucial predicate *onazi da* appears. When we have reached the end of the ambiguous region, we have in our mind a correspondence between Keynes and Natume Sooseki on the one hand and that between the circles Keynes was in and the circle of Sooseki's disciples on the other, correspondences which are based on linguistic features, that is, proper nouns in the former case and "noun *no atumari*" (*tisikizin no atumari* and *gakusya no atumari* vs. *desitati no atumari*) in the latter. Here we could perhaps speak of the principle of referential success (Crain and Steedman, 1985: 331)—very loosely construed, of course—being at work.<sup>30</sup> The principle is as follows:

If there is a reading that succeeds in referring to an entity already established in the hearer's mental model of the domain of discourse, then it is favored over one that does not.

Thus, the reading of (86) as '*Natume-Sooseki to [desitati no atumari]*' succeeds, as it were, in referring to Keynes and his circles already established in the reader's mental model by virtue of the formal correspondences explained above, and is favored over the reading '<*Natume-Sooseki to desitati*> *no atumari*' ('the gathering of Natume Sooseki and his disciples').

## 6. Evaluation

The plausibility of a hypothesized system of strategies is evaluated according to the degree to which it makes the right predictions, i.e., how much of the data it can account for. In order to see this, I have tested the system by applying the postulated strategies to all the relevant examples (300 in total) taken from four books (see Section 3). The result is that there are 272 regular examples and 28 problematic examples, the latter including: “problematic examples” discussed in Section 5; examples which allow more than one equally possible interpretation even in context. In addition, Yamanaka (1999) examined examples from a corpus of newspaper articles, all of which are of the “ $W_1$  no  $W_2$  no..... $W_m$  no X to Y no  $Z_1$  no  $Z_2$  no..... $Z_n$ ” type. The result is that 209 out of 219 examples are regular, with 10 problematic examples. In both cases, the success rate is high enough to be able to say that the system of strategies is a promising one.<sup>31</sup>

## 7. Conclusion

The resolution of ambiguities constitutes a major part of language understanding in general and sentence processing in particular. In order to deal with structural ambiguities in coordinate noun phrase structures in Japanese, a system of strategies based on conceptual information has been proposed, the application of which is guided by the integrated processing hypothesis. The system has been shown to apply to complex examples as well as simple ones, and corpus researches have confirmed its utility. In discussing complex examples, I have suggested that human sentence processing is flexible and that it is not always amenable to rigid algorithms often employed in computer-oriented frameworks.

Considering the wide range of examples the present system can handle, it can be said to be a good starting point, although some refinement may be necessary. The system should also be developed in wider scope. First, it is presently designed to deal with a particular type of structures. The problem remains whether similar systems of strategies can be proposed to resolve other types of structural ambiguities, and how they can be incorporated in a comprehensive model of sentence processing mechanisms. Second, in this paper the main focus has been on how several strategies interact to resolve ambiguities out of context. More effort is needed to elucidate in what exact process contextual information influences the application of the strategies in the ongoing understanding, either biasing a particular interpretation or canceling it. These two will be the main tasks of future research.

## Notes

- 1 Parts of this paper were presented at the Fifth Annual Meeting of the Association for Natural Language Processing, Tokyo, March 1999, the 118th General Meeting of the Linguistic Society of Japan, Tokyo, June 1999, and the 7th International Pragmatics Conference, Budapest, July 2000. I am grateful to many



scholars for valuable comments and suggestions on earlier versions of this paper.

- 2 Here the term *corpus* is used not in the currently popular sense of retrievable texts in the electronic medium but in the broader sense, that is, "a large collection of written or spoken texts that is used for language research" (Collins Cobuild English Dictionary, 3rd edn.).
- 3 Pickering et al. (2000: 19) hint at the possibility that there may be differences in language understanding mechanisms for different languages and for different syntactic constructions within languages. Crucially, Matsumoto (1997: Ch.3) claims that semantics and pragmatics play a much more prominent role than syntax in determining interpretations in Japanese, as far as noun-modifying constructions are concerned. For example, the following has a major reading in which the head noun *gakusei* is the missing subject of the predicate *katta*, but it can be an oblique case (from the student, for/to the student) depending on the context:  
[[*hon o katta*] *gakusei*]  
book ACC bought student  
'the student who bought a book' or 'the student from whom X bought a book' or 'the student for/to whom X bought a book'
- This is due to the absence of an explicit syntactic marker specifying the relation between the head noun and the clause, which also allows the most plausible locative interpretation of the head in the next example:  
[[*watasi ga kinoo tabeta*] *resutoran*]  
I NOM yesterday ate restaurant  
'the restaurant at which I ate yesterday'
- In contrast, interpretation of English relative clauses is typically guided by syntax, so that the counterpart of the above example *the restaurant that I ate yesterday* has only a pragmatically strange interpretation.
- 4 Taking the same example, Quirk et al. (1985: 960) say, "when coordinated heads are preceded by a determiner, the usual interpretation is that the determiner applies to each of the conjoins", but one of my informants says that the other interpretation that the determiner only applies to the first noun ('[his wife] and child') is not possible and the other informant says that it is only possible in an appropriate yet highly marked context. I am grateful to my former colleagues, Simon R. Potter and Neil J. Cowie, for information on this point and on the proper usage of some other English expressions.
- 5 Note that Japanese, being a SOV language, uses a postposition and always puts a modifier before the head noun, so that *onna no kekkon* means 'marriage of a woman'.
- 6 Lang (1984: 87ff.) claims that the sentence "John and Sue are married" is not ambiguous as between 'being married to one another' and 'being married each to someone else', but simply unspecified in these respects: the respective spouses of John and Sue are to be specified by contextual information available, and the former, combinatory interpretation occurs when Sue and John happen to be the spouses in question. While this argument sounds plausible with regard to symmetric predicates such as *be married*, a parallel explanation would seem to be somewhat far-fetched as to the noun phrase "salmon and cucumber sandwiches" to be discussed in Subsection 4.1. Although this problem merits further study, in this paper I regard the distinction between combinatory and segregatory interpretation as ambiguity.
- 7 Although the strategies I propose are not so comprehensive as many of those mentioned above, I prefer to call them *strategies* rather than *rules* in order to avoid the connotation of being all or nothing and that of conventionality as in "syntactic rules".
- 8 Following Cruse (2004: 219), I use the term *co-occurrence* in a broader sense than *collocation*. For further detail, see the discussion on the strategy based on co-occurrence preferences below.
- 9 There are a few types of apparent counterexamples regarding (i), but as Yamanaka (1988) shows, these can be taken care of by some independent pragmatic factors (see fn 26).
- 10 One might argue that the reason *watasi no kyoo* is strange is not semantic but syntactic: *kyoo* cannot take the head position of an NP because of its syntactic properties. It is true that a considerable number of Japanese lexical items have such properties, but *kyoo* does not seem to do so: we can naturally say *rainen no kyoo*, meaning 'today next year'.
- 11 These two articles may appear irrelevant here because they deal with relative clause attachment, but they are relevant to the overall structures considered in this paper. As the comments on (3) in Section 3 show, relative clauses are also included in the modifiers in question.
- 12 As mentioned in connection with (3), compound nouns in cited examples have no internal boundaries written in their original Japanese forms, but I insert a hyphen between their constituent nouns for the convenience of non-Japanese readers.
- 13 Nagao et al. (1983) also report that the existence of such words as *kankei* 'relation', *aida* 'interval', etc. has a biasing effect on the disambiguation of coordinated noun phrases as specified in Strategy (B), but they do not mention the inherent semantic feature of these words except that they are likely to be preceded by coordinated noun phrases.
- 14 Example (27) is the title of a letter to the editor and example (32) is the subtitle of a book. See the discussion at the beginning of Section 4 as to the interpretation of utterances out of context.

- 15 Quirk et al. (1985: 701) state that prepositional phrases denoting distinct semantic relations cannot be coordinated unless different processes are involved to give an effect similar to that of zeugma:  
 \*He was killed by a man and (by) an arrow.  
 (where *by a man* is 'agentive' and *by an arrow* is 'instrument'.)
- 16 In discussing the difference between *land* and *ground*, Fillmore (1982: 121) points out that *land* and *sea* on the one hand and *ground* and *air* on the other, each constitute a frame of its own. There is no reason, in my view, why *sea* and *sky* should not do so.
- 17 In English, there is a strong preference to attach a postmodifying relative clause to both conjuncts of a coordinate noun phrase structure (of the type *The lawyer greeted the powerful barrister and the wise judge who was/were walking to the courtroom*), as reported in the off-line written completion test by Schepman and Rodway (2000).
- 18 My terminology here is closely related to that of Cruse (1986): collocational anomaly corresponds to what he calls inappropriateness; semantic anomaly corresponds to paradox and incongruity combined; pragmatic anomaly corresponds to improbability. According to Cruse (2004: 220), inappropriateness results from the contravention of a collocational preference, whereas paradox and incongruity result from that of a selectional preference.
- 19 Regarding anomalous collocations such as 'the blond horse' and 'the bay boy', Lyons (1981: 161-162) states that even if a person's hair were of exactly the same color as the coat of a bay horse and if a horse's coat exactly matched the color of a blond person's hair, *blond* and *bay* would still not be used interchangeably. He concludes that "there are very many lexemes in all languages whose meaning cannot be considered to be totally independent of the collocations in which they most characteristically occur".
- 20 It may be that there is only one neon sign in the setting, with intermingled parts of red and yellow. If so, the correct interpretation is '[[[*mae no tatemono*] *no ikkai*] *no baa*] *no* [*aka to ki*] *no neon*']. The context does not make clear which is the case.
- 21 There are a number of studies relating to the range of information acquired during a fixation while reading a Japanese text. See Osaka (1998).
- 22 '*sihoo-seido ni okeru [kensatukan no sekinin]*' ('the prosecutor's responsibility in the judicial system') is the proper way of saying it.
- 23 In (59), Strategy (A-2) is also likely to apply, treating *tikara* and *gizyutu* as semantic correspondents. But because they are not quite typically in the relations illustrated in (37) through (45), I leave it an open question and give a decisive role to Strategy (A-1).
- 24 The additional use of slashes here has no special meaning other than to make clear which left-hand mark corresponds to which right-hand mark. Instead of using an extra representation (65), the problem regarding (64), i.e., the crucial effect that the possessive *Miura no* modifies *isi* 'will' and not *isi no tuyosa* 'strength of will' could also be accounted for as follows: if we assume an unexpressed possessive in front of *isi* as in *Miura no zinkeikan to Ø; isi no tuyosa*, then according to the strategy (7)-(i), the zero-possessive would first be attributed to *isi* alone. Such an analysis looks elegant but brings up another question: when and how is syntactic disambiguation based on an unexpressed element justified? Returning to the original problem, relevant are Staub (2007)'s findings that an initial syntactic analysis, even if it is transitory, remains present in the system during the period when an incompatible alternate analysis is dominant.
- 25 The additional use of braces here is motivated by the fact that *Z-D* and *Z-P* have different meanings in relation to *hon* from those they have in relation to the modifying clause.
- 26 I do not claim that syntactic error recovery never occurs in this type of modification. As argued in Yamanaka (1988), the following example seems to involve it:  
*Ookina ki no tukue*  
 big tree GEN desk  
 'a big wooden desk'
- Because *ookina ki* 'a big tree' is a normal co-occurrence, the default interpretation is '*[ookina ki] no tukue*'. But this interpretation is pragmatically rather strange, because it entails that the desk is made of wood which was part of a big tree, while we usually do not know or even think about the size of a tree which has provided wood for a desk. In other words, while we are interpreting *ki no tukue* 'a wooden desk', *ki* is taken as material, and as such it loses all the external characteristics (the size, the branches, etc.) that it formerly had when it was modified by *ookina*, so that a revised modification '*ookina [ki no tukue]*' is in order. This example differs from (66) in two respects: first, here it makes a substantial difference in meaning which noun the modifier *ookina* relates to; second, there is only a short distance between the modifier and the last noun, so that error recovery is not difficult.
- In this connection, based on their experiments on sentences involving NP/S ambiguities, Van Dyke and Lewis (2003) suggest that a longer distance between the head of the ambiguous region and the disambiguating word affects reanalysis because the items required for the correct analysis are subject to decay.

- 27 Based on their experiments on sentences involving NP/S ambiguities, Sturt et al. (2001) claim that people strongly prefer to avoid reanalysis if possible.
- 28 Regarding the problem discussed in (66) to (68), one could argue along the following lines that no extra mental representation is needed. That is, by taking the approach of a look-ahead buffer, which Hawkins (1994:61) generally postulates for head-final languages, one may say that in (66) the constituent modified by *watasi no yomitakatta* will be determined only at the end of the sequence, i.e., at *hon*. Now, if in (66) modification by *watasi no yomitakatta* is semantically compatible with either *Zyoan-Didian ya Zyoozi-Purinputon* or *hon* with little difference in meaning, the details of which constituent is syntactically modified can remain left open. I do not adopt such an analysis because this paper's principle is as early an attachment as possible and there is no reason to postpone decisions in (66). Interestingly, Christianson et al. (2001: 396ff.) suggest a possibility (which they say may seem radical from the perspective of current models of sentence comprehension) that, in the process of reanalyzing garden path sentences such as *While the man hunted the deer ran into the woods*, a sort of tree splicing might occur in which *the deer* serves as both object of the subordinate clause and subject of the matrix clause. Although their examples, unlike mine, show clear cases of misinterpretation, they also point to flexibility of human language understanding.
- 29 Here I intend the term *wider context* to include not only surrounding linguistic information but also non-linguistic information such as pictures and drawings, and cultural assumptions shared by writers and readers, but not knowledge immediately accessible through the expressions in the ambiguous region.
- 30 Altmann and Steedman (1988: 201) refined this principle as "the principle of referential support", which is more complex in content but more concise in formulation: an NP analysis which is referentially supported will be favored over one that is not. Although so called Referential Theory represented by the studies cited here involves a parallel model, it shares with my position the interactive view of human language understanding, which justifies my invocation of it here.
- 31 Given the description of the formal properties of the example data and the bibliographical information on the corpora in Section 3, it is possible for readers of Japanese to check the exactness of the success rates.

## References

- Altmann, Gerry, Steedman, Mark, 1988. Interaction with context during human sentence processing. *Cognition* 30, 191-238.
- Beaugrande, Robert-Alain de, Dressler, Wolfgang U., 1981. *Introduction to Text Linguistics*. Longman, New York.
- Bever, Thomas G., Sanz, Montserrat, Townsend, David J., 1998. The emperor's psycholinguistics. *Journal of Psycholinguistic Research* 27 (2), 261-284.
- Christianson, Kiel, Hollingworth, Andrew, Halliwell, John F., Ferreira, Fernanda, 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42 (4), 368-407.
- Crain, Stephen, Steedman, Mark, 1985. On not being led up the garden path: the use of context by the psychological syntax processor. In: Dowty, D. R., Karttunen, L., Zwicky, A. M. (Eds.), *Natural Language Parsing*. Cambridge University Press, Cambridge, pp. 320-358.
- Cruse, D. Alan, 1986. *Lexical Semantics*. Cambridge University Press, Cambridge.
- Cruse, D. Alan, 2004. *Meaning in Language* (2<sup>nd</sup> edn.). Oxford University Press, New York.
- Dik, Simon C., 1968. *Coordination*. North-Holland, Amsterdam.
- Fillmore, Charles J., 1982. Frame semantics. In: The Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm, Selected Papers from SICOL-1981*. Hanshin, Seoul, pp. 111-137.
- Ford, Marilyn, Bresnan, Joan, Kaplan, Ronald M., 1982. A competence-based theory of syntactic closure. In: Bresnan, J. (Ed.), *The Mental Representation of Grammatical Relations*. The MIT Press, Cambridge, MA, pp. 727-796.
- Frazier, Lyn, 1987. Sentence processing: a tutorial review. In: Coltheart, M. (Ed.), *Attention and Performance XII*. Erlbaum, Hove, pp. 559-586.
- Frazier, Lyn, Fodor, Janet D., 1978. The sausage machine: a new two-stage parsing model. *Cognition* 6, 291-325.
- Gibson, Edward, Loomis, Jacob, 1994. A corpus analysis of recency preference and predicate proximity. In: Ram, A., Eiselt, K. (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, Hillsdale, pp. 357-362.
- Gibson, Edward, Pearlmutter, Neal, Canseco-Gonzalez, Enriqueta, Hickok, Gregory, 1996. Recency preference in the human sentence processing mechanism. *Cognition* 59, 23-59.
- Hawkins, John A., 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press, Cambridge.
- Hirose, Yuki, 2006. Processing relative clauses in Japanese: coping with multiple ambiguities. In: Nakayama, M., Mazuka, R., and Shirai, Y. (Eds.), *The Handbook of East Asian Psycholinguistics Vol. 2: Japanese*. Cambridge University Press, Cambridge, pp. 264-269.

- Huddleston, Rodney, Pullum, Geoffrey K., 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge.
- Isiwata, Tosio, 1965. Heiritu-zyosi “ya” “to” no kinoo [The function of coordinate particles *ya* and *to*]. *Keiryoo-kokugogaku* [Mathematical linguistics] 32, 23-30.
- Isiwata, Tosio, 1968. Gengo no imi to gengo-zyoohoo-syori [Linguistic meaning and linguistic information processing]. In: Kokuritu-kokugo-kenkyuuzyo (Ed.), *Densi-keisanki ni yoru Kokugo-kenkyuu* [Computer-assisted Japanese language studies]. Syuuei-Syuppan, Tokyo, pp. 151-177.
- Kamide, Yuki, Mitchell, Don C., 1997. Relative clause attachment: nondeterminism in Japanese parsing. *Journal of Psycholinguistic Research* 26 (2), 247-254.
- Kaneko, Yasuo, 1987. Koozooyoo no ryooigi-bun no kaisyakuzoo no katayori to bun-syori-katei: nihongo no baai [Biases in the interpretations of structurally ambiguous sentences and the human sentence processing in Japanese]. *Kiso-sinrigaku-kenkyuu* [The Japanese journal of psychonomic science] 6 (1), 1-10.
- Kess, Joseph F., Nisimitu, Yoshihiro, 1989. *Linguistic Ambiguity in Natural Language*. Kuroosio-syuppan, Tokyo.
- Kimball, John, 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2 (1), 15-47.
- Kooij, Jan G., 1971. *Ambiguity in Natural Language*. North-Holland, Amsterdam.
- Lang, Ewald, 1984. *The Semantics of Coordination* (translation by J. Pheby). John Benjamins, Amsterdam.
- Lyons, John, 1981. *Language and Linguistics*. Cambridge University Press, Cambridge.
- Lytinen, Steven L., 1987. Integrating syntax and semantics. In: Nirenburg, S. (Ed.), *Machine Translation*. Cambridge University Press, Cambridge, pp. 302-316.
- MacDonald, Maryellen C., Pearlmutter, Neal J., Seidenberg, Mark S., 1994. Lexical nature of syntactic ambiguity resolution. *Psychological Review* 101 (4), 676-703.
- Mahesh, Kavi, Eiselt, Kurt P., Holbrook, Jennifer K., 1999. Sentence processing in understanding: interaction and integration of knowledge sources. In: Ram, A., Moorman, K. (Eds.), *Understanding Language Understanding*. The MIT Press, Cambridge, MA, pp. 27-72.
- Matsumoto, Yoshiko, 1997. *Noun-Modifying Constructions in Japanese*. John Benjamins, Amsterdam.
- Mazuka, Reiko, Itoh, Kenji, 1995. Can Japanese speakers be led down the garden path? In: Mazuka, R., Nagai, N. (Eds.), *Japanese Sentence Processing*. Lawrence Erlbaum Associates, Hillsdale, pp. 295-329.
- Miller, George A., Charles, Walter G., 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6 (1), 1-28.
- Miyamoto, Edson T., Gibson, Edward, Pearlmutter, Neal J., Aikawa, Takao, Miyagawa, Shigeru, 1999. A U-shaped relative clause attachment preference in Japanese. *Language and Cognitive Processes* 14 (5/6), 663-686.
- Mizutani, Sizuo, Tanaka, Sachiko, 1972. Go no heiretu-ketugooosi [On coordinate connectors of words or word groups]. *Keiryoo-kokugogaku* [Mathematical linguistics] 63, 19-36.
- Nagao, Makoto (Ed.), 1996. *Sizen-gengo-syori* [Natural language processing]. Iwanami-syoten, Tokyo.
- Nagao, Makoto, Tuzii, Zyun'iti, Tanaka, Nobuyosi, Isikawa, Masahiko, 1983. Kagaku-gizyutu-ronbun ni okeru heiretuku to sono kaiseki [Coordinate phrases in technological articles and their parsing]. *Zyooohoo-syori-gakkai-sizen-gengo-syori-kenkyuukai-siryoo* [Natural language processing workshop in the society of information processing] 36 (4), 1-9.
- Osaka, Naoyuki (Ed.), 1998. *Yomi* [Reading]. Asakura-syoten, Tokyo.
- Pickering, Martin J., Clifton, Charles, Jr., Crocker, Matthew W., 2000. Architectures and mechanisms in sentence comprehension. In: Crocker, M. W., Pickering, M., Clifton, C., Jr. (Eds.), *Architectures and Mechanisms for Language Processing*. Cambridge University Press, Cambridge, pp. 1-28.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, Svartvik, Jan, 1985. *A Comprehensive Grammar of the English Language*. Longman, New York.
- Sasano, Ryohei, Kawahara, Daisuke, Kurohashi, Sadao, 2005. Meisikaku-hureemu-zisyo no zidoo-kootiku to sore o motiita meisiku no kankei-kaiseki [Automatic construction of nominal case frames and its application to indirect anaphora resolution]. *Sizen-gengo-syori* [Natural language processing] 12 (3), 129-144.
- Schank, Roger, Birnbaum, Lawrence, 1984. Memory, meaning, and syntax. In: Bever, T. G., Carroll, J. M., Miller, L. A. (Eds.), *Talking Minds*. The MIT Press, Cambridge, MA, pp. 209-251.
- Schepman, Astrid, Rodway, Paul, 2000. Prosody and parsing in coordination structures. *The Quarterly Journal of Experimental Psychology* 53A (2), 377-396.
- Sperber, Dan, Wilson, Deirdre, 1995. *Relevance* (2nd ed.). Blackwell, Oxford.
- Staub, Adrian, 2007. The return of the repressed: abandoned parses facilitate syntactic reanalysis. *Journal of Memory and Language* 57, 299-323.
- Sturt, Patrick, Pickering, Martin J., Scheepers, Christoph, Crocker, Matthew W., 2001. The preservation of structure in language comprehension: is reanalysis the last resort? *Journal of Memory and Language* 45, 283-307.
- Van Dyke, Julie A., Lewis, Richard L., 2003. Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49, 285-316.

- Van Gompel, Roger P. G., Pickering, Martin J., Traxler, Matthew J., 2000. Unrestricted race: a new model of syntactic ambiguity resolution. In: Kennedy, A., Radach, R., Heller, D., Pynte, J. (Eds.), *Reading as a Perceptual Process*. Elsevier Science, Oxford, pp. 621-648.
- Wu, Haodong, Furugori, Teiji, 1998. A hybrid approach for resolving ambiguities in coordinate structures. *Sizen-gengo-syori* 5 (4), 3-16.
- Yamanaka, Nobuhiko, 1988. Nihongo no tagiteki na meisi-syuusyoku-koozoo no kaiseki [The disambiguation of “modifier noun *no* noun” construction in Japanese]. *Gengo-kenkyuu* [Linguistic researches] 94, 75-99.
- Yamanaka, Nobuhiko, 1999. Nihongo no tagiteki na heiretu-meisiku no kaiseki [The disambiguation of coordinate noun phrases in Japanese]. In: Gengo-syori-gakkai (Ed.), *Gengo-syori-gakkai Daigokai-nenzi-taikai Happyoo-ronbunshuu* [Proceedings of the fifth annual meeting of the association for natural language processing]. Gengo-syori-gakkai, Toyonaka, pp. 540-543.
- Yamanaka, Nobuhiko, 2000. Iwayuru “wakai otoko to onna” no tagisei ni tuite [On the so-called ambiguity of “young men and women”]. In: Yamada, S., Kikuti, Y., Momiyama, Y. (Eds.), *Nihongo Imi to Bunpoo no Huukai* [Japanese: the landscape of meaning and grammar]. Hituzi-syoboo, Tokyo, pp. 257-273.

## 概 要

現実の多義性であれ潜在的な多義性であれ、多義性の解消は、人間の言語理解の主要な部分を構成する。「東京のホテルとバーの数」(通常「[東京の [ホテルとバー]] の数」と解釈される)に見られるような、日本語の名詞句並列構造における構造的な多義性を解消するための仮説として、コーパスデータに基づいて3種類のストラテジーからなるインタラクティブなシステムを構築した。これらのストラテジーの適用は統合処理仮説によって導かれるものとする。提案したストラテジーの全ては、(A) 並列句同士の意味範囲の対応 (B) ある種の名詞に内在する「複数性」という概念 (C) 名詞句と修飾語句の間に存在する共起選好、といった概念的な情報に大部分基づいており、また、これらのストラテジーは基本的に修飾語句の早い付加を引き起こすべくデザインされている。これらのストラテジーからなるシステムは単純な構造を持つ例だけでなく複雑な構造を持つ例にも適用される。このシステムの有効性はコーパスの多くの用例によって確かめられた。複雑な構造を持つ例を論じた際に、人間による文処理は広く受け入れられている統語表示が想定するよりも柔軟性に富んでいることを示唆した。

**キーワード** 人間の言語理解、構造的な多義性、統合処理仮説、ストラテジー、修飾語句の早い付加