Diagram Understanding Using Integration of Layout Information and Textual Information

Yasuhiko Watanabe
Ryukoku University
Seta, Otsu
Shiga, Japan
watanabe@rins.ryukoku.ac.jp

Makoto Nagao
Kyoto University
Yoshida, Sakyo-ku
Kyoto, Japan
nagao@kuee.kyoto-u.ac.jp

1 Introduction

Pattern information and natural language information used together can complement and reinforce each other to enable more effective communication than can either medium alone (Feiner 91) (Nakamura 93). One of the good examples is a pictorial book of flora (**PBF**). In the PBF, readable explanations which combine texts, pictures, and diagrams are achieved, as shown in Figure 1 and 2. Taking advantage of this, we propose here a new method for analyzing the PBF diagrams by using natural language information.

In the PBF, many kinds of information about plants are generally stored in the following media:

- picture,
- explanation text, and
- diagram.

Pictures and explanation texts describe the shape, botany, and environment of each plant, as shown in Figure 1. On the contrary, as shown in Figure 2, diagrams describe very clearly the following kinds of information:

- correspondence between each part of the plant and its name
- part–whole relationship and synonymous words

As a result, diagrams are quite important in order to understand the explanation of each plant. For example, pictures and explanation texts cannot convey the information which is necessary to answer the following questions:

- which part of plant is *"takuyou* (stipules)"?
- what shape is *"hishinkei* (lanceolate shape)"?

Consequently, we have to investigate the method for understanding diagrams.

It is certain that diagrams are effective means of communication. However, it is also true that we feel difficulties when we try to understand diagrams without the explanation by natural language. From this, we conclude that natural language information is important to understand diagrams. Consequently, we propose a new framework for the in-
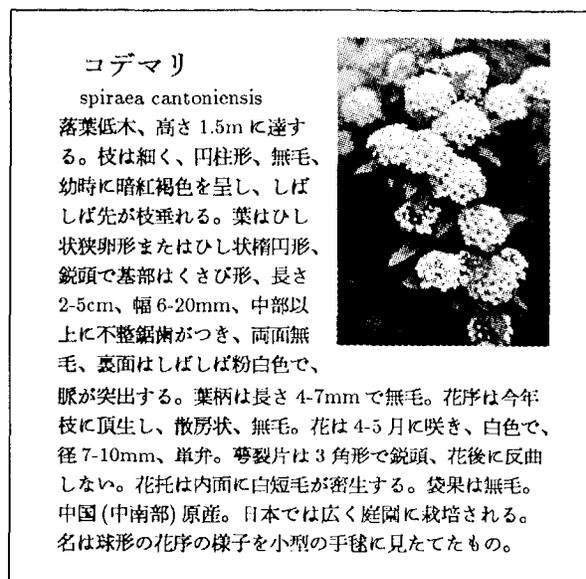

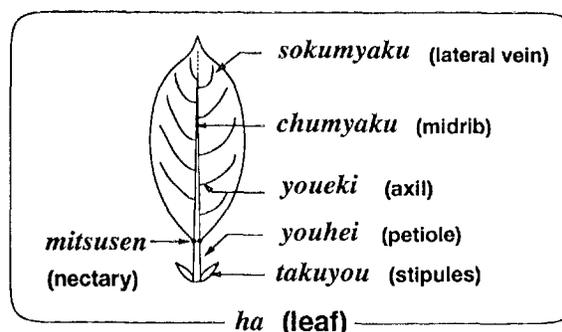Figure 1: An example of a PBF article (in Japanese)

Figure 2: An examples of PBF diagrams (leaf)

tegration of pattern (layout) information and natural language information for semantic understanding the PBF diagrams. In this study, for the observation and experiments, we use a PBF (in Japanese) the subject of which are wild flowers of Japan.
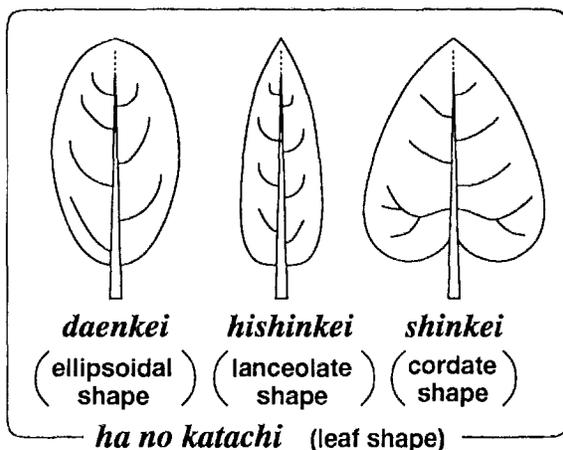
1374

daenkei (ellipsoidal shape) *hishinkei* (lanceolate shape) *shinkei* (cordate shape)

ha no katachi (leaf shape)

Figure 3: The shape of leaf

kajitsu (fruit) — *taika* (follicle) *'katsura'* *'natsumikan'* *mikanjyoka* (hespidium)

Figure 4: The variety of fruits

houhairekkai (loculicidal dehiscence)

shinpi no chuou ga sakeru (carpel splits open in its center)

rekkai (dehiscence)

Figure 5: The process of dehiscence

2 Diagram Understanding

2.1 PBF Diagram Understanding

A diagram in the PBF consists of its title and elements. A title shows the subject of the diagram and it is generally located under the diagram. Elements may be classified into three types:

- symbol,
- sketch,
- word.

Symbols (arrow, line, parenthesis, etc.) connect diagram elements. Sketches represent the features of a plant readably and accurately. The features represented by sketches are explained by words in the diagram. Observing the PBF diagrams, the information explained by words is classified into five categories, such as:

1. names of plant parts.
   (example) "*takuyou* (stipules)" in Figure 2

2. types of plant parts.
   (example) "*taika* (follicle)" and "*mikanjyoka* (hespidium)" in Figure 4

3. properties of plant parts.
   (example) "*daenkei* (ellipsoidal shape)" in Figure 3

4. names of plant species.
   (example) "*katsura*" and "*natsumikan*" in Figure 4

5. additional explanation.
   (example) "*shinpi no chuou ga sakeru* (carpel splits open in its center)" in Figure 5

Diagram understanding is the semantic interpretation of the elements in the diagram. As mentioned, in the PBF diagrams, the information represented by sketches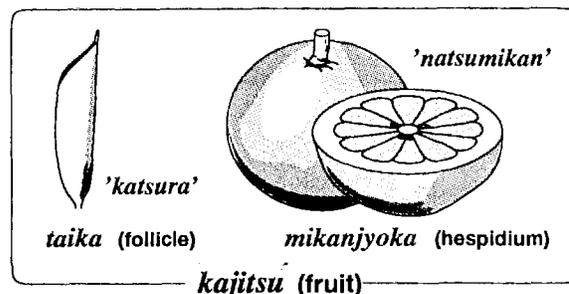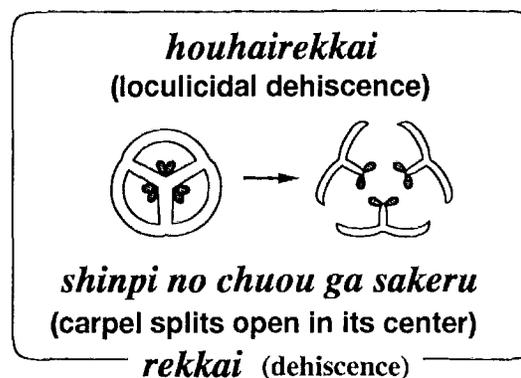 is explained by words. From this, we conclude that the semantic interpretation of the PBF diagrams is the classification of words in a diagram into these five categories.

For this purpose, we propose a new framework for the semantic interpretation of words in the PBF diagrams by using not only the pattern (layout) information but also the natural language information. The reason is as follows: there is no composing rule that strongly restricts the positions of elements and the semantic relations between elements. In other words, there are many ways of composing diagrams to explain an idea. For this reason, it is difficult to interpret the PBF diagrams only by using the pattern (layout) information. For example, "*daenkei* (ellipsoidal shape)", "*hishinkei* (lanceolate shape)", and "*shinkei* (cordate shape)" in Figure 3 represent the properties of the plant part, on the contrary, "*taika* (follicle)" and "*mikanjyoka* (hespidium)" in Figure 4 represent the types of the plant part. In spite of the semantic difference, all these words are located under the corresponding sketches, respectively.

2.2 Related Work

There are a few research topics related to diagram understanding. (Plant 89) (Futrelle 90) recognized the semantic structure of diagrams as the extension of diagram analysis. But they analyzed diagrams

by using knowledge about diagrams which is quite separate from natural language information. On the contrary, (Nakamura 93) analyzed diagrams in the encyclopedia dictionary by using its explanation texts and the thesaurus information. But it is difficult to analyze the PBF diagrams in the same way as (Nakamura 93) did. The reasons are as follows:

- It is certain that the explanation texts in the PBF are closely related to the PBF diagrams. However, these texts do not describe the contents of diagrams but the features of plants. That is, there is no explanation text for the PBF diagrams.

- Words in the PBF diagrams are generally technical terms which are not registered in the common thesaurus.

To solve these problems, we propose a new framework to analyze the PBF diagrams by using many kinds of clue expressions in the PBF explanation texts.

3 Information for PBF Diagram Understanding

3.1 Layout Information

For analyzing the PBF diagrams, we utilize two kinds of layout information. These are:

- type of relationships between diagram elements

- similarity of spatial relationships between diagram elements

Every word in the PBF diagrams is related to the other elements (sketches or other words). The relationships between words and the corresponding elements are classified into 2 types:

connection is the relationship between the word and its corresponding element. These are connected together by a symbol. For example, the relationship between "*takuyou* (stipules)" and the "leaf" sketch in Figure 2 is connection.

adjacency is the relationship between the word and its corresponding element. These are adjacent to each other and not connected by a symbol. For example, the relationship between "*natsumikan*" and the right sketch in the Figure 4 is adjacency.

A word connected by a symbol represents a name of the plant part. Consequently, in this case, the spatial relationship between the word and the corresponding element is not important for the semantic interpretation. For example, the semantic interpretation of "*mitsusen* (nectary)" in Figure 2 would remain unchanged even if "*mitsusen*" was located on the right of the "leaf" sketch.

On the contrary, a word which is not connected by a symbol may represent any type of information. Consequently, in this case, the spatial relationship between the word and the corresponding

element is important for the semantic interpretation. For example, it is inadequate to replace the position between "*mikanjyoka*" and "*natsumikan*" in Figure 4. It is because the replacement breaks the similarity of the spatial relationship which "*mikanjyoka* (hespidium)" and "*taika* (follicle)" have. In this way, words in the PBF diagrams which represent the same kind of information, often have the same spatial relationship. From this, we utilize the similarity of the spatial relationships for the propagation of the semantic interpretation in this way: suppose that words A and B have the same spatial relationship. If A is given the semantic interpretation but B is not, the semantic interpretation of A is given to B.

3.2 Natural Language Information

As mentioned previously, the PBF texts do not explain the PBF diagrams but describe many kinds of plants. The explanation texts, however, include many clues which are useful to classify the words in the PBF diagrams into the five semantic categories. In order to realize the semantic analysis, we utilize two kinds of natural language information for diagram understanding. These are:

- titles of the PBF articles.

- typical expressions which show implicitly the semantic interpretation of words in the PBF diagram.

Titles of the PBF articles represent names of the species. Typical expressions which we utilize for analyzing the diagrams are such as:

(a) A + *ha* + predicative noun

(b) A + *ga* + *aru* (exist)

(c) A + *ga* + verbalized noun + *suru*

(d) *ha* + A (A is a predicative noun)

(e) A (A is a verbalized noun) + *suru*

where A is a word in the diagram, "*ha*" and "*ga*" are Japanese postpositions, and "*suru*" accompanies a noun and verbalizes it. These five expression patterns are useful to interpret the words in the PBF diagrams. For example, words which represent the names of plant parts are found in expressions (a), (b), and (c), as shown in text (S–1) ~ (S–3), but not in expressions (d) and (e).

(S–1) *ha* (leaf) *ha daenkei* (ellipsoidal shape)

(S–2) *kibu ni* (at the base) *ha mitsusen* (nectary) *ga aru* (exist)

(S–3) *rimen ni* (in the back) *ha myaku* (vein) *ni sotte* (along) *ke* (hair) *ga missei* (glow tightly) *suru*

On the contrary, words which represent the properties are found in the expressions (d) and (e), as shown in texts (S–4) and (S–5), but not in expressions (a), (b), and (c).
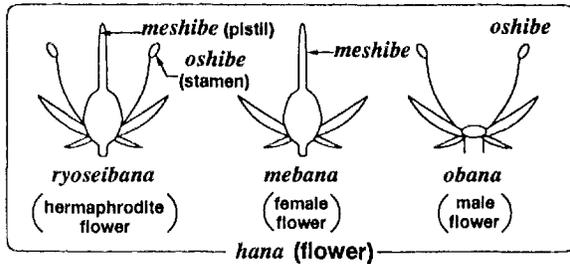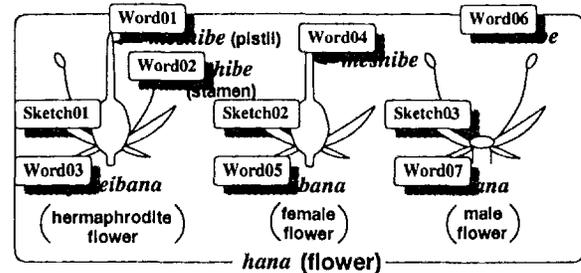
Figure 6: A diagram of flower

Figure 7: ID number of each word and sketch in Figure 6

ID Number	Writing	Corresponding element	type of correspondence	position
Word01	*meshibe*	Sketch01	connection	*
Word02	*oshibe*	Sketch01	connection	*
Word03	*ryoseibana*	Sketch01	adjacency	bottom
Word04	*meshibe*	Sketch02	connection	*
Word05	*mebana*	Sketch02	adjacency	bottom
Word06	*oshibe*	Sketch03	adjacency	above right
Word07	*obana*	Sketch03	adjacency	bottom

(a) layout information of Figure 6

Writing	Title	expression pattern				
		(a)	(b)	(c)	(d)	(e)
meshibe	0	46	23	0	0	0
oshibe	0	194	37	5	0	0
ryoseibana	0	4	0	0	1	0
mebana	0	26	2	0	1	0
obana	0	41	1	0	3	0

(Note)   expression pattern
(a) A + *ha* + predicative noun
(b) A + *ga* + *aru*
(c) A + *ga* + verbalized noun + *suru*
(d) *ha* + A (A is a predicative noun)
(e) A (A is a verbalized noun) + *suru*

(b) natural language information of Figure 6

Figure 8: An example of the layout and natural language information

(S–4) *kajitsu* (fruit) *ha kyukei* (spherical shape)
(S–5) *kajyo* (inflorescence) *ha tyousei* (terminal) *suru*

4   Process of PBF Diagram Understanding

4.1   Representation of Layout Information

Layout information is represented by hand in the following way.

Step 1. give ID number to all words and sketches in the diagram.

Step 2. describe the following kinds of information for each word in the diagram.

  1. ID number
  2. writing
  3. ID number of the corresponding element
  4. type of correspondence (connection or adjacency)

5. relative position from the center of the corresponding element when the type of correspondence is adjacency (we use 8 directions for the description: top, above right, right, below right, bottom, below left, left, above left)

Figure 7 shows the given ID number of each word and sketch in Figure 6. Figure 8 (a) represents the layout information of Figure 6. For example, Figure 8 (a) shows that "*ryoseibana* (hermaphrodite flower)" in Figure 6 has Word03 as the ID number, corresponds with Sketch01 (a sketch in the left of Figure 6), and is located under Sketch01. Moreover, by checking the information on the position, words which have the same spatial relationship can be detected. For example, Figure 8 (a) shows that there are words under the left sketch, under the central sketch, and in the above right and under the right sketch. Using this information, Word03 ("*ryoseibana*"), Word05 ("*mebana*"), and Word07

ID No.	Writing	Result	Rule
Word01	*meshibe*	name of plant parts	Rule 1
Word02	*oshibe*	name of plant parts	Rule 1
Word03	*ryoseibana*	type of plant parts	Rule 5
Word04	*meshibe*	name of plant parts	Rule 1
Word05	*mebana*	type of plant parts	Rule 5
Word06	*oshibe*	name of plant parts	Rule 4
Word07	*obana*	type of plant parts	Rule 5

Figure 9: Results of the semantic analysis for Figure 6

("*obana*") are detected as the words which have the same spatial relationship with the corresponding sketch.

4.2 Extraction and Representation of Natural Language Information

Natural language information, which is useful to interpret the words in the PBF diagrams, such as, titles and expression patterns, is extracted and represented in the following way:

Step 1. Extract titles from the explanation texts.

Step 2. Japanese morphological analysis. We used JUMAN(Kurohashi 97) as a Japanese morphological analyzer.

Step 3. Extract the following expression patterns by pattern matching from the results of the Japanese morphological analysis.

(a) A + *ha* + predicative noun

(b) A + *ga* + *aru* (exist)

(c) A + *ga* + verbalized noun + *suru*

(d) *ha* + A (A is a predicative noun)

(e) A (A is a verbalized noun) + *suru*

where A is a word in a diagram.

Step 4. Describe the results of Step 1 and 3 as the natural language information.

Figure 8 (b) shows the natural language information of Figure 6. The number of each expression pattern in Figure 8 (b) shows the frequency of it in the PBF explanation texts.

4.3 Semantic Analysis by Integration of Layout Information and Natural Language Information

In this section, we describe the process of semantic analysis for the PBF diagrams by using the integration of layout and natural language information. The semantic analysis is realized by applying the following rules in order: (Figure 9 shows the results of the semantic analysis for Figure 6)

Rule 1. [Rule for names of plant parts by using symbols]

A word which is connected to the other element by a symbol is interpreted as a name of the plant part. For example, "*meshibe*" (Word01 and Word04) and "*oshibe*" (Word02) in Figure 6, each of which is connected with its corresponding sketch by an arrow, are interpreted as the name of the plant part by this rule.

Rule 2. [Rule for names of plant species]

A word is interpreted as a name of the plant species, when it is:

(a) a title of the PBF articles, or

(b) written in Katakana letters [1].

For example, "*katsura*" and "*natsumikan*" in Figure 4 are interpreted as the name of the species by this rule. It should be noted that "*katsura*", that is a wild kind, is one in the titles of the PBF articles. On the contrary, "*natsumikan*", that is a cultivated kind, is not a title in the PBF. It is because the subject of the PBF which we used is wild flowers in Japan. As a result of this, "*katsura*" and "*natsumikan*" are interpreted by the condition above (a) and (b) in this rule, respectively.

Rule 3. [Rule for properties of plant parts]

A word in a diagram is interpreted as a property of the plant part when it is found in the expression pattern (d) and (e) described in Section 3.2, such as:

(d) *ha* + A (A is a predicative noun)

(e) A (A is a verbalized noun) + *suru*

and never found in the rest of the expression patterns in Section 3.2. For example, "*daenkei* (ellipsoidal shape)", "*hishinkei* (lanceolate shape)", and "*shinkei* (cordate shape)" in Figure 3 are interpreted as the properties of the plant part by this rule.

Rule 4. [Rule for names of plant parts by using the expression patterns]

A word in a diagram is interpreted as a name of the plant part when it is found in the expression patterns (a), (b), and (c) described in Section 3.2, such as:

(a) A + *ha* + predicative noun

(b) A + *ga* + *aru* (exist)

(c) A + *ga* + verbalized noun + *suru*

and never found in the rest of expression patterns in Section 3.2. For example, "*oshibe*" (Word06) in Figure 6 is interpreted as the name of the plant part by this rule.

---

[1] In Japanese PBF, names of species are generally written in Katakana letters. For example, "*katsura*" and "*natsumikan*" in Figure 4 are written in Katakana letters.

Table 1: Results of the semantic analysis

Relation	name of a plant part		property of a plant part		type of a plant part		name of species		additional explanation	
	success	failure	success	failure	success	failure	success	failure	success	failure
connection	74	0	—	—	—	—	—	—	—	—
adjacency	1	7	112	0	33	5	23	1	3	0

Rule 5. [Rule for types of plant parts]

Words in a diagram are interpreted as the types of the plant part when the following conditions are satisfied.

1. Each sketch is related to one of these words.
2. These words have the same spatial relationship with the corresponding sketch.
3. These words have the same Kanji character at the end of the writing.
4. Some of these words are found in the expression pattern (a), (b), and (c) described in Section 3.2:

   (a) A + *ha* + predicative noun
   (b) A + *ga* + *aru* (exist)
   (c) A + *ga* + verbalized noun + *suru*

For example, "*ryoseibana* (hermaphrodite flower)", "*mebana* (female flower)", and "*obana* (male flower)" in Figure 6 are interpreted as the types of the plant part by this rule.

Rule 6. [Propagation of semantic interpretation]

A word which cannot be interpreted by the rules 1–5 is given the same semantic information as the word which has the same spatial relationship. The way of detecting the words which have the same spatial relationship is described in Section 4.1.

**Rule 7.** [Rule for additional explanation]

An additional explanation generally includes

- postposition
- adjective
- verb
- adverb

Taking advantage of this, words are interpreted as an additional explanation when the result of their Japanese morphological analysis includes the above types of a part–of–speech. For example, "*shinpi no chuou ga sakeru* (carpel splits open in its center)" in Figure 5 is interpreted as an additional explanation by this rule.

5 Experimental Results

To evaluate our approach, we used 31 PBF diagrams in an experiment. These 31 diagrams included 175 sketches and 259 words. Table 1 shows
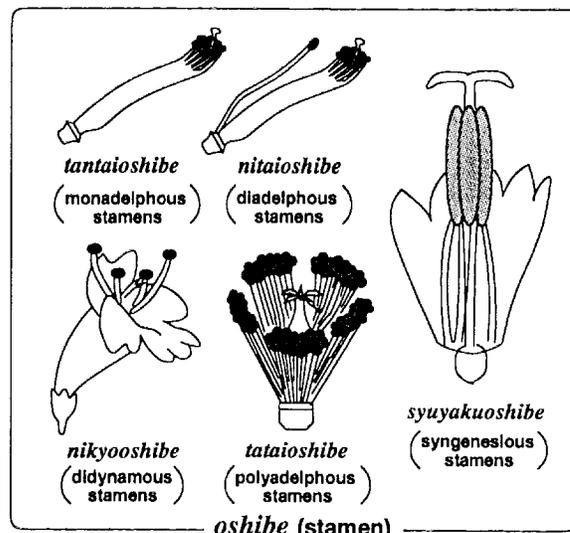
(a) A diagram of stamen
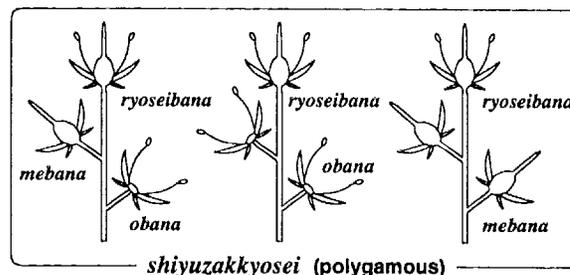
(b) A diagram of polygamous

Figure 10: Examples of incorrect analysis

the results of the semantic analysis of these diagrams.

Figure 10 gives two examples of the failures in this experiment. The words in Figure 10 (a) represent the types of "*oshibe* (stamen)", and could not be interpreted by our approach. This is because these words, such as "*nitaioshibe* (diadelphous stamens)", are rarely used in the PBF texts, and could not be found in five kinds of expression patterns in Section 3.2. The words in Figure 10 (b), "*ryoseibana* (hermaphrodite flower)", "*mebana* (female flower)", and "*obana* (male flower)" represent the names of the plant part. But these words could

not be interpreted by the rule 4 in our method because these words are found in the expressions, such as:

(S–6) *daibubun* (most of the flower) <u>*ha ryoseibana*</u> (hermaphrodite flower)

(S–7) *chuou no 1kko* (a flower in the center) <u>*ha*</u> <u>*mebana*</u> (female flower)

(S–8) *sokuhou no 2ko* (two flowers in the corner) <u>*ha obana*</u> (male flower)

6 Conclusion

At the moment, the pattern (layout) information is extracted and represented by hand. To realize an automatic extraction and representation of the pattern (layout) information, we have to investigate the following methods:

- a method for extracting the diagram elements

- a method for detecting the corresponding relations between the diagram elements

Fortunately, a large amount of diagrams is created and stored on computers. Taking advantage of this, we may avoid the difficulties in extracting the diagram elements by image processing. For this reason, we would like to investigate the method for detecting the spatial relationship between the diagram elements.

References

Feiner, McKeown: Automating the Generation of Coordinated Multimedia Explanations, IEEE Computer, Vol.24 No.10, (1991).

Futrelle: Strategies for diagram understanding: Generalized equivalence, spatial/object pyramids and animate vision, Proc. 10th ICPR, (1990)

Kurohashi, Nagao: JUMAN Manual version 3.4 (in Japanese), Nagao Lab., Kyoto University, (1997) [2].

Nagao: Methods of Image Pattern Recognition (in Japanese), CORONA, (1983).

Nakamura, Furukawa, Nagao: Diagram Understanding Utilizing Natural Language Text, 2nd International Conference on Document Analysis and Recognition, (1993).

Plant, Scrievner, Schappo, Woodcock: Usage and generality of knowledge in the interpretation of diagrams, Knowledge–Based Systems, Vol.2 No.2, (1989).

---

[2]The source file and the explanation (in Japanese) of Japanese morphological analyzer JUMAN can be obtained using anonymous FTP from
`ftp://pine.kuee.kyoto-u.ac.jp/pub/juman/juman3.4.tar.gz`