

# Long Sentence Analysis by Domain-Specific Pattern Grammar

Shinichi Doi, Kazunori Muraki, Shinichiro Kamei & Kiyoshi Yamabana

NEC Corp. C&C Information Technology Research Laboratories

4-1-1, Miyazaki, Miyamae-ku, Kawasaki 216, JAPAN

## 1 Long Sentence Analysis

We propose a method for analyzing long complex and compound sentences that utilizes global structure analysis with domain-specific pattern grammar.

Previously, long sentence analysis with global information used the following methods: two-level analysis—global structure analysis of long sentences with domain-independent function words and parsing of their constituents[Doi *et al.*, 1991], and pattern matching—adaptation of domain-specific fixed pattern to input sentences. By utilizing domain-dependent information the latter method could analyze long sentences of that domain. But since the matching is made only on the surface the sentence isn't analyzed well when patterns appear recursively.

## 2 Domain-Specific Pattern Grammar

Our method analyzes the global structure of long sentences by using three knowledge-bases: domain-specific patterns that can be described as a phrase structure grammar, a list of keywords that denote constituents of the patterns, and a pure basic grammar. An input sentence is initially parsed and divided into its constituents with these knowledge-bases, and then each constituent is parsed with a general grammar. Each constituent must be guaranteed uniformity by parsing with pure basic grammar.

To obtain a pattern grammar of Japanese long sentences we analyzed the structures of about 750 long sentences from the leads of news articles in a Japanese newspaper, Asahi Shinbun, and identified several fixed global patterns. An example of pattern grammar is shown in Fig. 1. Using the pattern grammar and keyword list(a-c), the global structure of the sentence(d) was analyzed as f).

## 3 Conclusion

Our method takes advantage of both two-level analysis and pattern matching, and can deal with the irregular appearance of patterns including recursive patterns, ellipsis of constituents and patterns that appear in only part of the sentence.

We have developed a Japanese lead analyzing system using a pattern grammar. We used this system with several 80-200 word Japanese leads such as the example sentence in Fig. 1., and obtained correct global structures and syntactic trees for them.

## References

[Doi *et al.*, 1991] Shinichi Doi, Kazunori Muraki, and Shinichiro Kamei. Lexical Discourse Grammar and its Application for Decision of Global Dependency (II). In *Proceedings WGNLC of the IE-ICE*, NLC91-29(PRU91-64), 1991. (in Japanese)

### Fig. 1 Example of Pattern Grammar and Keywords in Japanese

a) pattern grammar (*p.* denotes *phrase*)

statement report pattern

⇒ subject *p.* + date *p.* + situation *p.* + theme *p.* + direct statement *p.* + indirect statement *p.*

b) keyword list for a) (see e))

c) pure basic grammar noun *p.* ⇒ adnominal *p.* + noun *p.*, verb *p.* ⇒ adverbial *p.* + verb *p.*

d) example sentence of Japanese lead

Nakasone-shushou-wa 17-nichi, shuuin-honkaigi-de okonawareta kakutou-daihyou-shitsumon-deno touben-de, Kokka-himitsu-houan-nitsuite "kokueki-wo mamoru-tame-nimo nanrakano datouna sochi-ga hitsuyou"-to nobe, jimintou-giin-niyoru houan-sai-teishutsu-ni maemuki-no shisei-wo shimeshita.

e) keywords in d) (*mk. pp.* denotes *marking postposition*)

-wa(subject *mk. pp.*) / 17-nichi(day), / touben(answer)-de(location *mk. pp.*), / -nitsuite(theme *mk. pp.*) / -to(quotation *mk. pp.*) nobe(said), / shisei(attitude)-wo(object *mk. pp.*) shimeshita(showed).

f) global structure of d)

subject *p.* : Nakasone-shushou-wa (Prime Minister Nakasone)

date *p.* : 17-nichi, (on 17th.)

situation *p.* : shuuin-honkaigi-de okonawareta kakutou-daihyou-shitsumon-deno touben-de, (in the answer to the party leaders' interpellations at the plenary session of the House of Representatives)

theme *p.* : Kokka-himitsu-houan-nitsuite (about the National-Secret-Bill)

direct : "kokueki-wo mamoru-tame-nimo nanrakano datouna sochi-ga hitsuyou"-to nobe, statement *p.* (said "some appropriate measures are needed to defend the national interest")

indirect : jimintou-giin-niyoru houan-sai-teishutsu-ni maemuki-no shisei-wo shimeshita.

statement *p.* (showed a positive attitude toward re-introducing of the bill by the Liberal Democrats)