The treatment of Japanese focus particles based on Lexical-Functional Grammar

Tomoko OHKUMA
Corporate Research Group
Fuji Xerox Co., Ltd.
430 Sakai, Nakai, Ashigarakamigun
Kanagawa, Japan
Ohkuma.Tomoko@fujixerox.co.jp

Hiroshi MASUICHI
Corporate Research Group
Fuji Xerox Co., Ltd.
430 Sakai, Nakai, Ashigarakamigun
Kanagawa, Japan
Hiroshi.Masuichi@fujixerox.co.jp

Hiroki YOSHIMURA
Corporate Research Group
Fuji Xerox Co., Ltd.
430 Sakai, Nakai, Ashigarakamigun
Kanagawa, Japan
Hiroki.Yoshimura@fujixerox.co.jp

Yasunari HARADA
School of Law, Waseda University
1-6-1 Nishi-waseda, Shinjuku,
Tokyo, Japan
harada@waseda.jp

Abstract

The purpose of this paper is to discuss treatment of 'focus particles' in Japanese sentences to be incorporated in a Japanese parsing system based on the Lexical-Functional Grammar(LFG) formalism. Focus particle can follow nouns, quantifiers, verbs, other particles (postpositions) and auxiliary verbs. Thus, it is necessary for a large-scale grammar to treat focus particles properly. Furthermore, there are syntactic and semantic ambiguities caused by the particles. We propose phrase structure rules and lexical entry constraints which cover focus particles in various positions and account for the ambiguities.

1 Introduction

As has been observed in many studies, various functional relations are expressed by particles in Japanese. In particular, particles such as 'kurai', 'bakari', 'dake', 'nomi', 'hodo' and 'made' specify focus in sentences. It has been observed that these particles mark contrastive focus within sentences (Numata et al., 2000). It is important for a Japanese parsing system to be able to treat these focus particles properly, because focus particles bear different syntactic functions depending on where they appear in a given Japanese construction. In this paper, we propose phrase structure rules and lexical entry constraints to treat Japanese focus particles on the basis of the Lexical-Functional Grammar (LFG) formalism (Kaplan and Bresnan, 1982; Dalrymple, 2001). The proposed rules and constraints for treating focus particles have been incorporated into a Japanese LFG system we are currently developing and have been used extensively for parsing real-world Japanese text.

2 Japanese LFG system

Our Japanese LFG parsing system development aims at broad coverage and deep analysis (Masuichi and Ohkuma, 2003; Masuichi et al., 2003).

The lexical entries for the words in the input are created by the output of the Chasen morphological analyzer (Matsumoto et al., 1999). In addition to these lexical entries, we have other lexicon files for verbs, adjectives, adjectival nouns and function words.

Lexical entries for verbs, adjectives and adjectival nouns were written based on the case frame information in the Japanese IPAL dictionary (IPA, 1987) and have been manually enhanced.

Lexical entries for verbs consist of 10,387 entries and 41,115 functional annotations for 2,366 verbs. Lexical entries for adjectives and for adjectival nouns consist of 947 entries and 2,197 functional annotations for 369 words in total.

Lexical entries of focus particles are included in our 'core lexical entries'. Core lexical entries include those for basic words such as auxiliary verbs, postpositional particles and so forth, plus syntactically important nouns such as 'toki' (time) and 'aida' (interval) and consist of 1,252 entries and 1,913 functional annotations for 675 words.

Phrase structure rules and constraints for focus particles are described in the grammar rules. The grammar rules include 2,468 terms in their disjunctive normal forms and 1,223 functional annotations. The Japanese grammar rules have been designed in such a waty that we could capture cross-linguistic parallelisms (Ohkuma and Masuichi, 2002; Kim et al., 2003), which is one of the major objectives of the Parallel Grammar Project (parGram) (Butt et al., 2002). ParGram uses the parser and grammar development platform XLE for six languages: English, French, German, Norwegian, Japanese, and Urdu .

XLE outputs all possible analyses of the input as c(onstituent)-structures and f(unctional)-structures. C-structure encodes phrasal dominance and precedence relations and is represented as a phrase structure tree. F-structure encodes syntactic predicate argument structure and is represented as an attribute-value matrix (Maxwell and Kaplan, 1993).

3 Japanese focus particles

Japanese focus particles can follow nouns, quantifiers, verbs, auxiliary verbs and other particles (Numata et al., 2000). We assume they are basically classified into the four types shown in (A)-(D).

(A) After nouns or quantifiers

> (1) *kare-dake-ga Tokyo-e   ik-ta.*
> he.F.nom   Tokyo.to go.past
> Only he went to Tokyo.

> (2) *Kaiin-wa   zyosei-nomi da.*
> member.T.nom woman.F   is.
> The members are only women.

> (3) *kanozyo-wa 3ko-hodo ringo-o   taber-ta.*
> she.Top.nom 3.parts.F apple.acc eat.past
> She ate about 3 apples.

(B) After verbs (sentential clauses)

> (4) *naku-kurai kitui sigoto da-ta.*
> cry.F   hard work   was.
> It was such hard work that I almost cried.

> (5) *Tyotto   miru-dake desu.*
> little   look.F   is.
> Just looking.

(C) After auxiliary verbs

> (6) *kanozyo-wa Tyokoreto-o   taber-te-bakari-iru.*
> she.T.nom   chocolate.acc eat-particle.F.Prog.
> She has been just eating chocolates.

449

(7) *oo   sugite   moter-nai-kurai-da.*
 much   too   hold.neg.F

 It is so much that someone almost can't hold it.

(D) After case-marking particles

(8) *Kare-ni-dake-wa kono nyusu-o   tutaer-tai.*
 he.dat.F.T   this   news.acc tell.Want

 (I) want to tell this news only to him.

(9) *kare-wa   Gakko-de-bakari hon-o   yomu.*
 he.Topic.nom school.Obl.F   book.acc read

 He read a book just in school.

This behavior of the focus particles indicates that they have similar syntactic distribution to adverbs in English. They also have similar meanings to adverbs in English. For example, 'dake' and 'nomi' can be translated as 'only' and 'hodo' as 'approximately'. The following phrase structure rules (10)-(13) simply treat focus particles as adverbs (ADJUNCTs) as in (Butt et al., 1999). Rule (10) corresponds to (A), (11) to (B), (12) to (C) and (13) to (D). These rules cover the examples in (A)-(D).

(10) N $\longrightarrow$ N   FP.
            ($\uparrow$=$\downarrow$) ($\uparrow$ADJUNCT=$\downarrow$)

(11) S $\longrightarrow$ S   FP.
            ($\uparrow$=$\downarrow$) ($\uparrow$ADJUNCT=$\downarrow$)

(12) AUX $\longrightarrow$ AUX*   FP*
                  ($\uparrow$GF* ADJUNCT=$\downarrow$)

(13) PP $\longrightarrow$ PP*   FP*
                 ($\uparrow$GF* ADJUNCT=$\downarrow$)

4 Ambiguity of focus particles

The rules above cannot treat the ambiguity of the sentence in (14) in which the focus particle 'kurai' follows the pronoun 'kare' (he).

(14) *kare-kurai eigo-o   benkyosur-ta-darou.*
  he.F   English.acc   study.Past-may

There are two possible interpretations of (14), as given in (15) and (16). Rule (10) induces the parsing result that corresponds to (16) only, and the interpretation (15) is missing.

(15) He may have studied English if nobody else does.

(16) Someone may have studied English as hard as he does.

The same problem occurs with the focus particle 'made'. Sentence (17) has two possible interpretations: (18) and (19). Rule (10) induces a parsing result only for (19).

(17) *yusyoku-made nobir-ta.*
  dinner.F   postponed.Passve.Past

(18) Even the dinner was postponed.

(19) Something was postponed until the dinner.

450

5 Phrase structure rules for focus particle

We account for the phenomena shown in the examples above by taking the position that Japanese focus particles have two distinct syntactic functions (Mizutani, 1990). The focus particles in (A) and (B) are categorized as a kind of suffix (Focus Suffix, FS) that follows nouns and verbs, while the focus particles in (C) and (D) are categorized as a kind of postposition (Focus PostPosition, FPP) that follows particles and auxiliary verbs. A phrase together with a following FS functions as an adverb. Therefore, we postulate the phrase structure rule and lexical entry constraint shown in (20).

(20) ADV $\longrightarrow$ { Sadj | Nadj}   FS.
　　　　　　　　　　($\uparrow$=$\downarrow$)

　　FS kurai   ($\uparrow$FOCUS-FORM)='kurai'.

Rules for FPP are shown in (21) and (22).

(21) NP $\longrightarrow$ Nadj   {ϵ | PP | FPP}*.
　　　　　　　　　($\uparrow$=$\downarrow$)

　　FPP kurai   ($\uparrow$FOCUS-FORM)='kurai'.

(22) VP $\longrightarrow$ V   {PP | AUXverb | FPP}*.
　　　　　　　　　($\uparrow$=$\downarrow$)

　　FPP kurai   ($\uparrow$FOCUS-FORM)='kurai'.

In the analysis of sentences like (14), both rule (20) and rule (21) are used (invoked).

When (20) is applied, our Japanese LFG system outputs the c-structure shown in Fig. 1 and the corresponding f-structure shown in Fig. 2. This analysis corresponds to the interpretation in (15). In this analysis, 'kare' (he) is the SUBJ in (14); the system interprets the focus particle 'kurai' as an FPP that follows a dropped nominative case marker (ϵ).
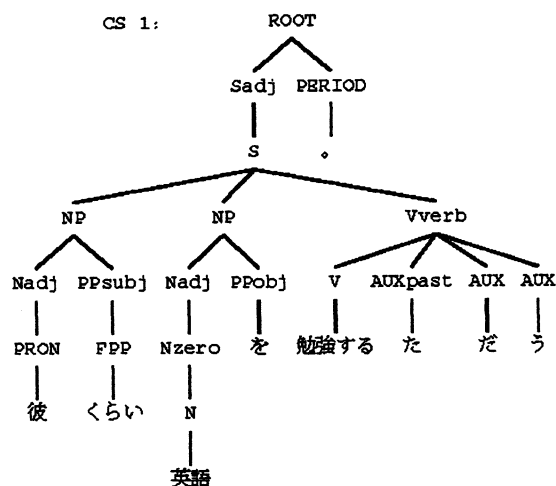
Figure 1: c-structure for (15)

451

"彼 くらい 英語 を 勉強する た だ う 。"

```
┌PRED      '勉強する <[0:pro], [5:英語] >'                                                                              ┐
│         ┌PRED 'pro'                                                                                               │
│SUBJ     0│ANIM +, CASE nom, FOCUS-FORM 'くらい', GEND-SEM male, NUM sg, PERS 3, PRON-FORM 彼, PRON-TYPE pers│
│         ┌PRED '英語'            ┐                                                                                 │
│OBJ      5│CASE acc, PERS 3│                                                                                      │
│ADDRESS  {presumption}                                                                                            │
│DP        [FOCUSnominal { [0:pro]}]                                                                               │
│TNS-ASP   [MOOD indicative, TENSE past]                                                                           │
│11PASSIVE -, STMT-TYPE decl, VTYPE main                                                                           │
└                                                                                                                  ┘
```

Figure 2: f-structure for (15)

When (21) is applied, our system outputs the c-structure in Fig. 3 and the corresponding f-structure shown in Fig. 4. This analysis corresponds to the interpretation in (16). In this analysis, 'kare' is an ADJUNCT in (14); the system interprets the focus particle 'kurai' as FS that follows the pronoun 'kare' as a suffix. Sentence (17) is analyzed in the same way as (14). The above rules for Japanese focus particles can output two possible analyses for sentences like (14) and (17).

Note that the syntactic structures represented in the c-structures shown in Fig. 1 and Fig. 3 are almost the same. The semantic difference between (15) and (16) is realized in the f-structures shown in Fig. 2 and Fig. 4.
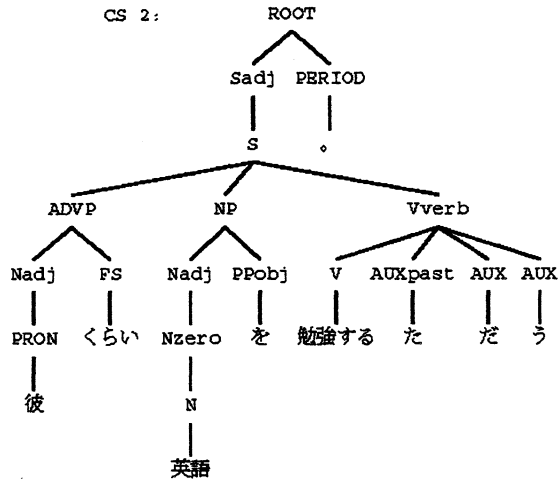
```
CS 2:              ROOT
                   /\
               Sadj  PERIOD
                |     |
                S     。
        _____|_____
       |         |               |
     ADVP        NP            Vverb
     /\          /\          /  |  \  \
  Nadj  FS    Nadj  PPobj   V AUXpast AUX AUX
   |    |      |     |       |   |    |   |
  PRON くらい  Nzero を    勉強する た   だ  う
   |          |
   彼          N
              |
             英語
```

Figure 3: c-structure for (16)

"彼 くらい 英語 を 勉強する た だ う 。"

```
┌PRED      '勉強する <[11-SUBJ:pro], [5:英語] >'                                                                      ┐
│         ┌PRED      'pro' ┐                                                                                        │
│SUBJ     │PRON-TYPE null│                                                                                         │
│         ┌PRED '英語'           ┐                                                                                 │
│OBJ      5│CASE acc, PERS 3│                                                                                      │
│         ┌PRED 'pro'                                                                                              │
│ADJUNCT  0│ADV-TYPE vp-adv, ANIM +, FOCUS-FORM 'くらい', GEND-SEM male, NUM sg, PERS 3, PRON-FORM 彼, PRON-TYPE pers│
│ADDRESS  {presumption}                                                                                            │
│TNS-ASP  [MOOD indicative, TENSE past]                                                                            │
│11PASSIVE -, STMT-TYPE decl, VTYPE main                                                                           │
└                                                                                                                  ┘
```

Figure 4: f-structure for (16)

452

Furthermore, (20)-(22) do not overgenerate analyses from unambiguous input sentences, while the rules account for the ambiguities caused by focus particles. When our system parses (3), (21) and (22) do not apply. Thus, our system outputs only one analysis by using (20); this is shown in Fig. 5 and Fig. 6. In other words, 'hodo' in (3) is interpreted as FS, but not FPP, in our analysis.
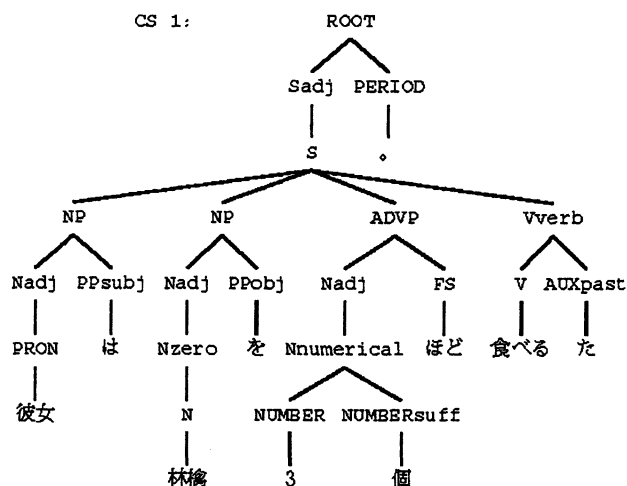
CS 1:

```
                      ROOT
                      /\
                Sadj   PERIOD
                 |      |
                 S      。
        _____/|_____
       NP      NP    ADVP    Vverb
      /\      /\     /\       /\
  Nadj PPsubj Nadj PPobj Nadj  FS   V   AUXpast
   |    |    |    |    |    |   |     |
  PRON  は  Nzero を Nnumerical ほど 食べる  た
   |         |      /\
  彼女       N   NUMBER NUMBERsuff
             |      |      |
            林檎     3      個
```

Figure 5: c-structure for (3)

"彼女 は 林檎 を 3 個 ほど 食べる た 。"

```
┌PRED   '食べる <[0:pro], [8:林檎] >'
│SUBJ    ┌PRED 'pro'
│       0│ANIM +, CASE nom, GEND-SEM female, NUM sg, PERS 3, PRON-FORM 彼女, PRON-TYPE pers, TOPICALIZATION-FORM'は'
│OBJ     ┌PRED '林檎'
│       8│CASE acc, PERS 3
│ADJUNCT ┌          ┌PRED 'pro'
│       {│          │NTYPE [GRAIN pronoun]
│        │          │SPEC [NUMBER 14[PRED '3' CLASSIFIER-FORM'個', NUMBER-TYPE card]]
│        │        20│ADV-TYPE vp-adv, FOCUS-FORM'ほど', PRON-TYPE 'null'
│DF      [TOPICnominal { [0:pro]}]
│TNS-ASP [MOOD indicative, TENSE past]
│      24 PASSIVE -, STMT-TYPE decl, VTYPE main
```
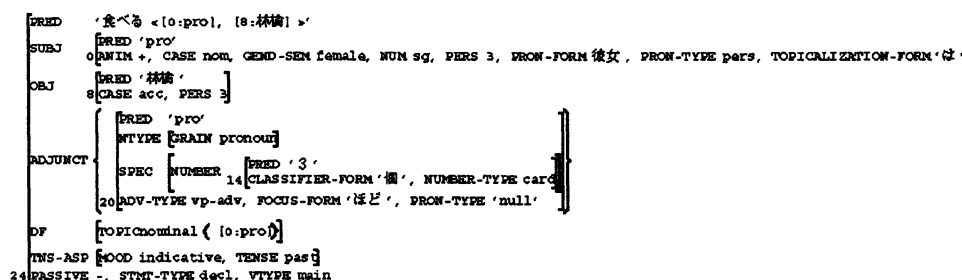
Figure 6: f-structure for (3)

Our rules do not apply (20) and (21) when parsing (6). Thus, our system outputs only one analysis by using (22); this is shown in Fig. 7 and Fig. 8. In other words, 'bakari' in (6) is interpreted as FPP, but not FS, in our system.
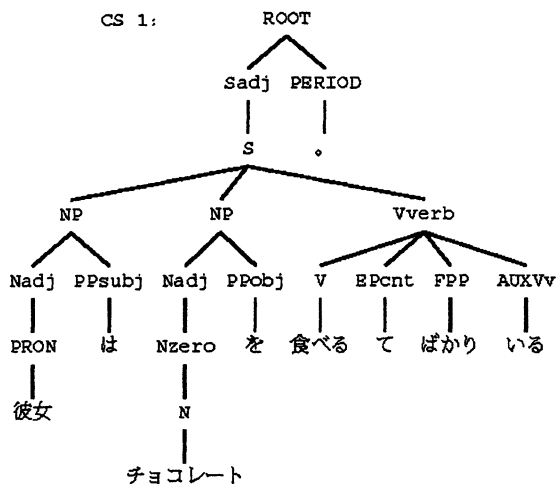
CS 1:   ROOT
                         /\
                    Sadj   PERIOD
                      | |
                      S   ｡
              /‾‾‾‾‾‾‾/\‾‾‾‾‾‾‾\
            NP   NP   Vverb
           /\   /\   /‾|‾\
       Nadj PPsubj Nadj PPobj V EPcnt FPP AUXVv
        | | | | | | | |
      PRON は   Nzero を   食べる て ばかり いる
        | |
      彼女   N
                        |
                    チョコレート

Figure 7: c-structure for (6)

"彼女 は チョコレート を 食べる て ばかり いる 。"

```
PRED    '食べる <[o:pro], [8:チョコレート] >'
SUBJ    [PRED 'pro'
        [ANIM +, CASE nom, GEND-SEM female, NUM sg, PERS 3, PRON-FORM 彼女, PRON-TYPE pers, TOPICALIZATION-FORM 'は'
OBJ     [PRED 'チョコレート'
        [CASE acc, PERS 3 ]
DF      [FOCUSverbial [DFOBJ [8:チョコレート]
                      [DFPRED 食べる
        [TOPICnominal < [o:pro]>
TNS-ASP [MOOD indicative, PROG +, TENSE pred]
FOCUS-FORM 'ばかり ', PASSIVE -, STMT-TYPE decl, VTYPE main
```

Figure 8: f-structure for (6)

6 Conclusion

In this paper, we proposed phrase structure rules and lexical entry constraints that successfully treat Japanese focus particles on the basis of the LFG formalism. The proposed rules and constraints cover the numerous constructions with focus particles, while properly accounting for ambiguities caused by focus particles and not producing redundant analyses. We incorporated the proposed rules and constraints into our Japanese LFG system and obtained reasonable c-structures and f-structures for sentences with focus particles.

References

Butt, M., Dyvik H., T.H. King, H. Masuichi, and C. Rohrer. 2002. The parallel grammar project. In *COLING 2002 Workshop on Grammar Engineering and Evaluation*, pages 1–7.

Butt, M., T.H. King, M.-E. Niño, and F. Segond. 1999. *A Grammar Writer's Cookbook*. CSLI Publications, Stanford.

Dalrymple, M. 2001. *Lexical Functional Grammar*. Academic Press, New York.

IPA. 1987. Lexicon of the japanese language for computers ipal. Technical report, Information-technology Promotion Agency'(IPA).

Kaplan, R.M. and J. Bresnan. 1982. *Lexical-Functional Grammar: A formal system for grammatical representation*. The MIT Press, Cambridge, MA.

Kim, R., M. Dalrymple, R.M. Kaplan, T.H. King, H. Masuichi, and T. Ohkuma. 2003. Multilingual grammar development via grammar porting. In *2003 European Summer School in Logic, Language, and Information (ESSLLI03) workshop "Ideas and Strategies for Multilingual Grammar Development"*. To appear.

Masuichi, H. and T. Ohkuma. 2003. Constructing a practical Japanese parser based on Lexical Functional Grammar. *Journal of Natural Language Processing (in Japanese)*, 10(2):79–109.

Masuichi, H., T. Ohkuma, H. Yoshimura, and Y. Harada. 2003. Japanese parser on the basis of the Lexical-Functional Grammar formalism and its evaluation. In *The 17th Pacific Asia Conference on Language, Information and Computation*. To appear.

Matsumoto, Y., A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, K. Takaoka, and M. Asahara. 1999. Japanese morphological analysis system Chasen version 2.0 manual 2nd edition. Technical report, Nara Institute of Science and Technology.

Maxwell, J. T, III. and R.M. Kaplan. 1993. The interface between phrasal and functional constraints. *Computational Linguistics*, 19:571–589.

Mizutani, S. 1990. *Notes of Japanese grammar*. Kohan press (in Japanese), Tokyo.

Numata, Y., S. Kinsui, and M. Kudo, 2000. *Tence, Negation, Focus*, pages 153–216. Iwanami (in Japanese), Tokyo.

Ohkuma, T. and H. Masuichi. 2002. Treatment of omitted postpositionals and auxiliary verbs. *IPSJ SIG Notes (in Japanese)*, 150:141–148.