PARSING JAPANESE HONORIFICS IN UNIFICATION-BASED GRAMMAR

Hiroyuki MAEDA, Susumu KATO, Kiyoshi KOGURE and Hitoshi IIDA

ATR Interpreting Telephony Research Laboratories
Twin 21 Bldg. MID Tower, 2-1-61 Shiromi, Higashi-ku, Osaka 540, Japan

Abstract

This paper presents a unification-based approach to Japanese honorifics based on a version of HPSG (Head-driven Phrase Structure Grammar)[1][2]. Utterance parsing is based on lexical specifications of each lexical item, including honorifics, and a few general PSG rules using a parser capable of unifying cyclic feature structures. It is shown that the possible word orders of Japanese honorific predicate constituents can be automatically deduced in the proposed framework without independently specifying them. Discourse Information Change Rules (DICRs) that allow resolving a class of anaphors in honorific contexts are also formulated.

1. Introduction

Japanese has a rich grammaticalized system of honorifics to express the speaker's honorific attitudes toward discourse agents (i.e. persons who are related to the discourse). As opposed to such written texts as scientific or newspaper articles, where the author's rather 'neutral' honorific attitude is required, in spoken dialogues, an abundant number of honorific expressions is used and plays an important role in resolving human zero-anaphors.

In this paper, a unification-based approach to Japanese honorifics is proposed. First, Mizutani's theory of honorific expression act[3] is introduced to define basic honorific attitude types used in specifying pragmatic constraints on the use of Japanese honorifics. Then a range of honorifics are classified into subtypes from a morphological and syntactico-semantic perspective and examples of their lexical specifications are shown. The main characteristics of the utterance parser and an approach to explaining possible word orders of honorific predicate constituents are described. Finally, Discourse Information Change Rules are formulated that resolve a class of anaphors in honorific contexts.

2. Speaker's Honorific Attitudes toward Discourse Agents

2.1. Grammatical Aspects of Honorifics

A distinction must be made between the speaker's honorific attitude as determined by the utterance situation (the social relationship between discourse agents, the atmosphere of the setting, etc), and the honorific attitude as expressed by special linguistic means independent of the utterance situation. For example, by violating a usage principle for the determination of an honorific attitude (i.e. "one should not exalt oneself in front of others"), uses of an honorific expression about the speaker himself can function as a kind of joke. However, without the help of grammatical properties of honorifics independent of particular utterance situations, the violation of a usage principle itself could not be recognized at all, thus the expression could not function as a joke. Though the former situational determination of honorific attitude is an interesting subject matter for socio and psycho-linguistic researchers, the latter grammatical properties of honorifics are our concern here and what is described with lexical specifications for honorifics.

2.2. Mizutani's Theory of Honorific Expression Act

Mizutani's theory of honorific expression act is introduced to define basic honorific attitude types that stipulate the pragmatic constraints on Japanese honorifics. In this model, discourse agents are positioned in an abstract two-dimensional honorific space (Fig. 1). How they are positioned is a socio and psycho-linguistic problem, which is not pursued here.
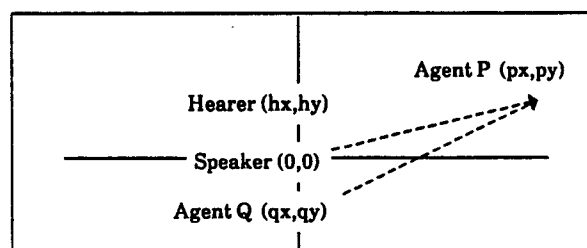

Fig 1. Honorific Space

An honorific expresson act reflects the configuraion of these discourse agent points. The speaker is set as the point of origin, and the speaker's honorific attitude toward a discourse agent, say P, is defined as the position vector of point P. The speaker's honorific attitude toward agent P relative to agent Q is defined as a vector from point Q to point P. The value and the direction of the vector are defined as follows:

Honorific Value :

for $v = (x. y)$, the honorific value of a vector v (written as $|v|$) is defined as:

$$|v| = \begin{cases} y & \text{iff } x = 0; \\ 0 & \text{iff } x \neq 0; \end{cases}$$

Honorific Direction :

a. *up* $|v| > 0$,
b. *down* $|v| < 0$,
c. *flat* $|v| = 0$ and $x = 0$,
d. *across* $|v| = 0$ and $x \neq 0$.

[N.B.]   Assuming an honorific space to be two dimensional (not one dimensional), an *across* direction can be distinguished from a *flat* direction. An *across* direction of a vector corresponds to the case where no positive honorific relation between the two agents (i.e. *up*, *down*, or *flat*) is recognized by the speaker.

Though the speaker's honorific attitudes can be characterized from several viewpoints (e.g. *up/down, distant/close, formal/informal*), Mizutani's model is appropriate for describing Japanese honorifics because the *up/down* aspect most relevantly characterizes Japanese honorifics. Moreover, it is not clear how the other aspects are independently grammaticalized in the Japanese honorific system.

Based on the direction of the vector defined above, the following four subtypes of honorific attitude relations are distinguished.

Honorific Attitude Type :

a. *honor-up*
b. *honor-down*
c. *honor-flat*
e. *honor-across*

3. Description of Japanese Honorifics

3.1. Classification of Japanese Honorifics

3.1.1. Morphological Viewpoint

In Japanese, words in a wide range of syntactic categories (i.e. nouns, verbs, adjectives, nominal-verbs, nominal-adjectives, etc) are systematically put into their honorific forms. They are classified into two subtypes according to how they are derived from their nonhonorific forms.

Classification by the lexical derivation type:

honorific-word =
   a. *regular-form-honorific-word*
     (e.g. "o-kak-i" from "kak-i" [write$_{vinf}$])
     [HP-[write$_{vstem}$-CS$_{inf}$]]
   b. *irregular-form-honorific-word*
     (e.g. "ossyar-" from "iw-" [speak$_{vstem}$])

[N.B.]   HP and CS stand for 'Honoric Prefix' and 'Conjugation Suffix' respectively. Words is transcribed in its phonemic representation.

While regular-form honorific words share a common base with their nonhonorific forms because they are derived by the productive honorific-affixation process, irregular-form honorific words have special word forms that have no direct connection to their nonhonorific forms. This distiction plays an important role in the lexical specification of honorifics and in possible word orders of Japanese honorific predicate constituents.

3.1.2. Syntactico-Semantic Viewpoint

In traditional school grammar, Japanese honorifics have been classified into three categories: respect words ('sonkeigo'), condescending words ('kenjougo'), and polite words ('teineigo'). However, in this traditional tripartite classification, common features of respect-words and condescending-words not shared by polite-words are not explicit. That is, while an agent toward whom the speaker's honorific attitude is expressed must be grammatically located in the sentence (i.e. as subject or object) in the case of respect or condescending words, this requirement does not apply to polite words. Thus a more elaborate classification is adopted. Conventional terms are replaced by Harada[4]'s more syntactico-semantically motivated ones.

Classification by the syntactic role of an agent to whom the speaker's honorific attitude is expressed:

honorific-word =
   a. *propositional-honorific-word* =
     a.1. *subject-honorific-word* (respect-word)
       (e.g. "kudasar-u" [give$_{vsenf}$])
     a.2. *object-honorific-word* (condescending-word)
       (e.g. "sasiage-ru" [give$_{vsenf}$])
   b. *performative-honorific-word* (polite-word)
     (e.g. 'des-u', 'mas-u')

[N.B.]   For example, a verb which takes a nonanimate subject (e.g. "fur-u" in the sentece "Ame (rain) ga (SBJ) fur-u(fall)." [The rain falls.]) can be put into its performative honorific form ("Ame ga fur-i mas-u."), but not into its subject honorific form (* "Ame ga o-fur-i ni nar-u."). This is in accordance with the difference between propositional honorifics and performative honorifics.

[N.B.]   There are a class of words which function in between the a.2 and b types of honorifics (e.g. "mair-u" [go/come$_{vsenf}$] in "Basu ga mair-i mas-u." [A bus will come.]). Let us call them *propositional-performative-words*.

Minus-honorifics are given no place in the traditional tripartite classification. However, they are classified in our approach as correponding to the expressed honorific attitude types.

140

honorific-word =

    a. *plus-honorific-word*
        (e.g. "aw-a-re-ru" [meet$_{regular-sbjhon}$])
            [[...meet$_{vstem}$-CS$_{vong}$]-PlusHonAux$_{vstem}$-CS$_{senf}$]
    b. *minus-honorific-word*
        (e.g. "aw-i-yagar-u" [meet$_{regular-sbjhon}$])
            [[...meet$_{vstem}$-CS$_{inf}$]-MinusHonAux$_{vstem}$-CS$_{senf}$]

[N.B.] The Japanese honorific system has no systematized means to positively express *honor-flat* or *honor-across* honorific attitudes. An non-honorific plain word form may express *honor-flat* honorific attitudes toward a discourse agent in a situation such as speaking to an old friend, while it may express *honor-across* honorific attitudes in a situation such as writing a technical paper.

Because the classfications of honorifics from different viewpoints as summarized above are cross-categorical, and thus independent of one another, a single honorific word (e.g. "hozak-u" [say$_{vsenf}$]) can function at the same time as *irregular-form-honorific-word, subject-honorific-word,* and *minus-honorific-word.*

3.2. A Unification-based Lexical Approach

A unification-based lexicalism approach is adopted here for describing Japanese honorifics for the following reasons:
(a) a unification-based approach enables the integrated description of information from various kinds of sources (syntax, semantics, etc), thus allowing their simultaneous analysis;
(b) a lexical approach helps to increase the modularity of grammar. In this approach, a grammar has only a small number of general syntactic rule schemata and most of grammatical information is to be specified in a lexicon. Linguistic word-class generalizations can be formed by making grammatical categories complex by representing them with feature-structures.

The specification of verbal category honorifics is important because the verbal categories are the most productive in the honorification process, and thus appropriate to clearly show how diverse aspects of the Japanese honorific system are described in this approach.

3.3. Examples of lexical specifications

3.3.1. Regular-Form Honorifics

Subject Honorification by "V$_{vong}$ + (ra)re-ru"

Regular form honorifics are compositionally analyzed by giving lexical specifications for each honorific-word formation formative. For example, most plain-form verbs can be put into their simple subject-plus-honorific form by postpositioning the auxiliary verb "(ra)re-ru" to them ("re-ru" and "rare-ru" are allomorphs of a single morpheme). Lexical information for these formatives is specified in the feature structure:

```
[[orth(orthography) ?orth]
 [head [[pos(part-of-speech) v]
        [ctype(conjugation-type) vowel]
        [cform(conjugation-form) stem]]]
 [adjacent ?pred]
 [subcat {
     ?sbj[[head [[pos p]
                [grf(grammatical-function) sbj]]]
          [subcat {}]
          [sem ?sbjsem]
          [senf [[human +]]]]
     ?pred[[head [[pos v]
                 [ctype ?predctype]
                 [cform vong(voice-negative)]]
           [subcat {?sbj}]
           [sem ?predsem]]}]
 [sem ?predsem]
 [prag [[restrs {[[reln honor-up]
                 [origin *speaker*]
                 [goal ?sbjsem]]}]]]])

where <?orth ?predctype> ∈ {<"re"    cons>
                            <"rare*  (:or vowel
                                          kuru
                                          suru)>}
```

Fig 2. Lexical Specification for a simple subject-plus honorification morpheme ("(ra)re-ru")

[N.B.] ? is a prefix for a tag-name used to represent a token identity of feature-structures. *Speaker* is a special global variable bound to a feature structure representing the speaker's information.

The 'prag' feature describes the pragmatic constraint on this expression (the "honor-up" relationship from the speaker to the subject agent of the predicate is required for this expression to be used in a pragmatically appropriate way). Description with the 'honor-up' honorific attitude relation shows that this expression is a 'plus-honorific' expression. Structure-sharing of the 'goal' feature value of this honorifc attitude relation with the semantic value of the predicate's subject shows that this expression is a 'subject-honorific' expression. The requirement for the 'orth' feature value (?orth) and the 'ctype' value in the 'subcat' feature (?predctype) describes the morphophonemic characteristic of this morpheme by stipulating that 're-(ru)' subcategorize for either a regular consonant-stem ctype verb or an irregular ctype verb ('suru'[do]), and that 'rare-(ru)' subcategorize for either a regular vowel-stem ctype verb or an irregular ctype verb ('kuru' [come]), correctly allowing (1a) and (1c) but not (1b).

(1) a. *Sensei ga kyoositu*   e   *ika*   <u>re</u>   *ta.*
    teacher SBJ classroom   to   go[ctype vowel]   Past
    "(The) teacher went to (the) classromm."

    b. * *Sensei ga kyoositu*   e   *ika*   <u>rare</u> *ta.*

    c. *Sensei ga kyoositu*   e   *ko*   <u>rare</u> *ta.*
                          come[ctype kuru]
    "(The) teacher came to (the) classroom."

    d. * *Kyoositu*   e   *ko*   *Sensei ga* <u>rare</u>   *ta.*

141

The 'adjacent' feature is a special feature which assures that its value be the first element in the list when the set description in the 'subcat' value is expanded into list descriptions by a rule reader. The specification of this feature implies that this morph is a bound morph and thus requires its adjacent element to be realized as a nonnull phonetic form. Though the set description in the 'subcat' value is introduced to allow word order variation among complement daughters in Japanese, without this kind of specification, ungrammatical sequences such as (1d) are also allowed for auxiliariy verbs.

[N.B.] A set description in the subcat feature of a feature sturucture,[[adjacent ?c][subcat {?a ?b ?c}]], for example, is expanded into its corresponding two possible list descriptions by a rule reader as follows: [[adjacent ?c][subcat (:or <?c ?b ?a> <?c ?a ?b>)]. Furthermore, <?c ?b ?a>, for example, is expanded into a feature structure such as [[first ?c][rest [[first ?b][rest [[first ?a][rest end]]].

Object Honorification by "HP + V_{inf} + suru"

Next, let us consider a more complicated formation pattern for deriving a regular object-plus-honorific form. As productive as the above "V_{vong} + (ra)re-ru" pattern is, an "HP + V_{inf} + suru" pattern can put most verbs with two grammatical human arguments into their corresponding object honorific forms as follows:

"o + aw-i + suru" from "aw-" (meet_{vstem}),
"go + shoukai + suru" from "shoukai" (introduce_{vstem}).

[N.B.] "o-" and "go-" are two forms of a single morpheme (honorific prefix) that is prefixed to words in a variety of syntactic categories (See Appendix I). The choice depends on the following element's origin. If the element is a Sino-Japanese morpheme (kango), the honorifc prefix takes the form "go-"; if it is a native one, the honorific prefix is realized as "o-", though there are exceptions.

In a naive analysis of Japanese honorifics, these honorific forms derive from their corresponding plain forms by a simple object honorification lexical rule that does not take into account their internal constituent structures (e.g. "aw-u" → "o-aw-i-suru"). Accordingly, this kind of naive analysis is inadequate for the following reasons:
(a) it is arguable that "HP + V_{inf}" forms a unit in some structural level before forming the unit "HP + V_{inf} + suru", considering the existence of such constructions as "HP + V_{inf} + ni + nar-u" (normal-sbj-plus-hon-form), "HP + V_{inf} + negaw-u(request)", and "HP + V_{inf} + itadak-u(receive-favor_{irregular-obj-plus-hon-form})", but this assertion is not explicitly illustrated in a naive analysis;
(b) though some adverbial postpositions such as "wa" (contrastive), "mo" (also) and "sae" (even) can appear inside the object honorific form (e.g. "o-aw-i-WA-suru", "go-shoukai-SAE-MO-suru"), it is difficult to derive these forms by a naive analysis in light of the generalization concerning

adverbial postpositions appearing in other environments (e.g. "Sensei ga kyoositu DAKE e WA ko rare ta" [the teacher came only to the classroom]);
(c) a naive analysis fails to explain the kind of the elements that can operate as a V_{inf} element in the pattern, which is automatically explained in the proposed framework as will be shown in section 5.

This regular object-plus-honorification process is compositionally analyzed in the proposed framework by giving each of its formatives a lexical specification, in the same manner as the "V_{vong} + (ra)re-ru" pattern subject-plus-honorific analysis.

Here the expression "o-aw-i-suru" is analized. Fig 3.a represents the lexical information of the verb "aw-" (meet) in its infinitive form ("aw-i").

```
[[orth "aw-i"]
 [can-take-hp +][lex +]
 [head [[pos v][ctype cons][cform inf]
        [hpform "o"]]]
 [subcat {[[head [[pos p][grf sbj][form ga]
                   [semf [[human +]]]]]
          [subcat ()]
          [sem ?sbjsem]]
         [[head [[pos p][grf obj][form ni]
                  [semf [[human +]]]]]
          [subcat ()]
          [sem ?objsem]]}]
 [sem [[reln meet]
       [agent ?sbjsem]
       [object ?objsem]]]]
```

Fig 3.a. Lexical Information for "aw-i" (meet_{vinf})

First, honorific prefixation lexical rule is applied to this infinitive-form verb. Fig 3.b represents the lexical information of an honorific prefix (HP) and Fig 3.c shows how this lexical rule is stated in the proposed framework.

```
[[orth ?hpform]
 [head [[pos hp]
        [coh [[can-take-hp +][lex +]
              [head [[pos v][cform inf]
                     [hpform ?hpform]]]]]]]
 [subcat ()]]
```

Fig 3.b. Lexical information for HP preceding Vinf

```
(defrule x -> (hp x)
  (<0 can-take-hp> == -)
  (<1 head coh> == <2>)
  (<0 head> == <2 head>)
  (<0 subcat> == <2 subcat>)
  (<0 sem> == <2 sem>)
  (<0 prag restrs> == (:union <1 prag restrs>
                              <2 prag restrs>)))
```

Fig 3.c. Honorific prefixation rule

[N.B.] The rule stated in an extended version of PATR-II notation consists of two parts; CFG-part and constraints. CFG-part is used to propose an efficient top-down expectation in the parser. Constraints are required for the rule application to end successfully. Here, all constraints are described by equations of two feature structures. "< >" is used to denote a feature structure path, and "==" to denote a token identity relation between two feature structures.

The 'head|coh(CategoryOfHead)' feature of a category specifies the kind of its head. An HP can take a lexical infinitive-form verb whose 'can-take-hp' value is ' + '. An HP is assigned its appropriate realization form[*] (in this case, "o" form), because its 'orth' value and the head's 'hpform' value are the same. The first equation in the rule statement prevents a second application of the honorific prefixation rule to the same verb (*"o-o-aw-i") by specifying that the mother category's 'can-take-hp' feature value be '-'.[**] The other equations in the rule are ones common to the adjunct-head structures.

[*N.B.] A note is needed here concerning the realization of HP. When the adjacent feature of the second right-hand-side symbol in the CFG-part is nil as in the above case, it is enough just to concatenate both 'orth' feature values of the right-hand-side symbols and make it the 'orth' feature value of the left-hand-side symbol. However, when the head element's adjacent feature has a nonnull value (i.e. in the case that the head element is a bound morph), a more complicted operation is needed. But here we only mention its necessity and avoid its precise formulation to save space.
[**N.B.] The 'can-take-hp' feature is specified as '-' not only for already HP-prefixed elements, but also for almost all irregular form honorific verbs (e.g. *"o-osshar-i"[say], *"o-itadak-i"[receive-favor]) and most mono-syllabic infinitive-form verbs that have corresponding irregular-form honorifics (e.g. *"o-si" [do], *"o-mi" [look at]).

Next, the usual complement-head structure rule (Fig 3.d) is applied to the resulting feature structure for "o-aw-i" and the feature structure for a normal object-plus honorification formative ("-suru", as shown in Fig 3.e). Thus the normal object plus honorifc form ("o-aw-i-(suru)") for "aw-"[meet] is obtained in a compositional way.

```
(defrule m -> (c h)
  (<0 head> == <2 head>)
  (<1> == (:first <2 subcat>)
  (<0 subcat> == (:rest <2 subcat>))
  (<0 sem> == <2 sem>)
  (<0 prag restrs> (:union <1 prag restrs>
                           <2 prag restrs>)))
```

Fig 3.d. Complement head structure rule

```
[[orth ""]
 [head [[pos v][ctype suru][cform stem]
        [irregular-cforms [[vong si][inf si]···]]]]
 [can-take-hp -]
 [adjacent ?pred]
 [subcat {?sbj[[head [[pos p][grf sbj]
                      [semf [[human +]]]]]]
          [subcat {}]
          [sem ?sbjsem]]
          ?obj[[head [[pos p][grf obj]
                      [semf [[human +]]]]]]
          [subcat {}]
          [sem ?objsem]]
          ?pred[[head [[pos v][cform inf][hp +]]]
          [subcat {?sbj ?obj}]
          [sem ?predsem]]}]
 [sem ?predsem]
 [prag [[restrs {[[reln honor-up]
                  [origin ?sbjsem]
                  [goal ?objsem]]}]]]]]
```

Fig 3.e. Lexical Specification for a normal object-plus honorification formative ("(-suru)")

3.3.2. Irregular Form Honorifics

Irregular form honorifics share most of their lexical information with their nonhonorific counterparts. In our framework, redundant lexical specification for irregular-form honorifics is avoided by using lexical inheritance mechanism from their superclasses. For example, the necessary lexical specification for the irregular subject honorific form "(-te)itadak-" of the donatory auxiliary verb "(-te)moraw-" is reduced, as shown in Fig 4.a. This turns out to be equivalent to Fig 4.b by unifying pieces of information from its super-classes, *te-receive-favor* and *obj-plus-hon*.

```
(:superclasses te-receive-favor obj-plus-hon)
[[orth "itadak"]
 [head [[ctype cons][cform stem]]]])
```

Fig 4.a. Neccesary lexical specification for the irregular form donatory auxiliary verb"(-te)itadak-"

```
[[orth "itadak"]
 [head [[pos v][ctype cons][cform stem]]]
 [subcat {[[head [[pos p][grf sbj][form ga]]]
           [subcat {}]
           [sem ?sbjsem]]
          [[head [[pos p][grf obj][form ni]]]
           [subcat {}]
           [sem ?objsem]]
          [[head [[pos v][cform te]]]
           [subcat {[[head [[pos p][grf sbj]]]
                     [subcat {}]
                     [sem ?objsem]]}]
           [sem ?predsem]]}]
 [sem [[reln transfer-favor]
       [donator ?sbjsem]
       [donatee ?objsem]
       [accompanied-action ?predsem]]]
 [prag [[restrs {[[reln honor-up]
                  [origin ?sbjsem]
                  [goal ?objsem]]
                 [reln empathy-degree]
                 [more ?sbjsem]
                 [less ?objsem]]}]]]]])
```

Fig 4.b. Whole lexical Information for "(-te)itadak-"

Lexical Information for other irregular-form honorifics is likewise specified.

4. Unification-based CFG Parser

Fig 5 shows the organization of the unification-based CFG parser. The parser is essentially based on Earley's algorithm, and unifies feature structures in its completion process. The description of grammatical rules and lexical items are complied into feature structures by the rule reader.

Unification of cyclic feature structuers might be necessary to analyze certain expressions. To give some examples:
(a) frozen honorific words such as "o-naka" (belly) and "go-ran" (to look at) must always be prefixed by an HP (the element in bold face);
(b) the polite form ("gozar-") of the verb "ar-"/"ir-" (to be) almost always needs to be followed by the polite honorific auxiliary verb "-masu" in modern Japanese.
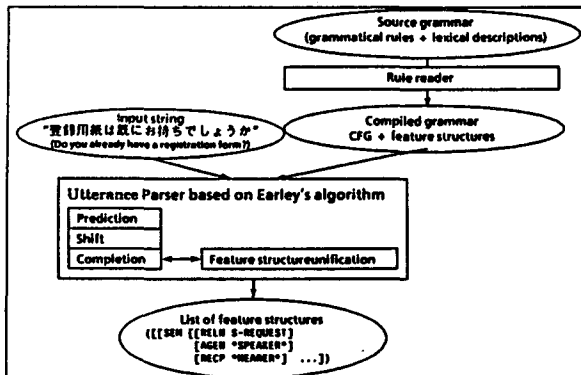
Fig 5. Organization of the Unification-based Parser

In describing the above linguistic phenemena, it is convenient if requirements for its head category can be specified not only for adjunct elements, but also for complement elements. In such cases, one more equation as follows needs to be added to the usual head-complement structure rule statement shown in Fig 3.d.

<1 head coh> = = <2>

The complied feature structure for the equations in Fig 3.d plus the above equation includes a cyclic structure as shown in Fig 6

An extended version of Wroblewski[5]'s feature structure unification algorithm was developed to allow rule statements including cycles[6]. The extended algorithm can unify cyclic feature structures while avoiding unnecessary overcopying of feature stuructures.

5. Word Order of Honorific Predicate Constituents

In Japanese, a verbal predicate is composed of one main verb and postpositioned auxiliary verbs (though possibly none exist). Because both main verbs and auxiliary verbs may have honorific forms, various sequences of honorifics might be expected to occur in a predicate as a simple matter of possible combinations. However, their possible word orders are restricted by a grammatical principles. Traditionally, possibile word orders were described in detail and the
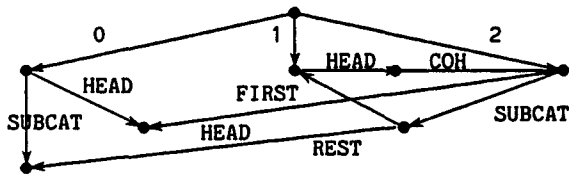

Fig 6. Cyclic part of the compiled feature structure

explanations for them were given from a rather speculative perspective. In this research, it is shown how possible word orders can be deduced from lexical specifications of honorifics.

5.1. Propositional and Performative Honorifics

A propositional honorific formative always precedes a performative honorific formative. For example, though "awa-re-masu" ([[[meet$_{vong}$]-SbjPlusHon]-PerformativeHon]) and "o-awi-si-masu" ([[[HP-meet$_{vinf}$]-ObjPlusHon]-PerformativeHon]) are possible expressions, they would be impossible if their word orders were reversed (i.e. performative honorific placed before propositional honorific).

This restriction on word order is considered a consequence of the lexical specifications for both types of honorifics. As shown in section 3, propositional honorification formatives subcategorize a verbal category whose subject (and object) elements are not filled yet as its adjacent element. On the other hand, a performative honorification formative subcategorizes a verbal category with saturated subcategorization. This represents the lexical specification for "masu".

```
[[orth ""]
 [head [[pos v][ctype masu][cform stem]
        [irregular-cforms [[senf masu]···]]]]
 [can-take-hp -]
 [adjacent ?pred]
 [subcat {?pred[[head [[pos v][cform masu]]]
               [subcat {}]
               [sem ?predsem]]}J
 [sem ?predsem]
 [prag [[restrs {[[reln honor-up]
               [origin °speaker°]
               [goal °hearer°]]}]]]]]
```

Fig 7. Lexical Specification for a performative honorification formative "masu"

The performative honorificaton formative "masu" cannot, therefore, immediately precede a propositional honorification formative due to the requirement concerning the adjacent element of propositional honorifics. The opposite order, however, constitutes a syntactically legitimate structure.

5.2. Subject and Object Honorifics

An object honorific formative must precede a subject honorific formative, though there is an important class of exceptions (verbs that subcategorize a 'te' form verb as an adjacent element such as "(-te)itadak-"[receive-favor]). For example, "o-awi-sa-reru" ([[[HP-meet$_{vinf}$]-ObjPlusHon]-SbjPlusHon]) is a possible word order, but "o-awa-re-suru" ([[HP-[meetvong-SbjPlusHon]]-ObjPlusHon]) is not possible if "-re(ru)" is used as an honorification formative. This word order restriction can be explained in the same way as for the

144

above case: that is, as shown in section 3, the normal object honorification formative "-suru" subcategorizes a verb whose subject and object are not yet filled. The simple subject honorification formative "-(ra)reru" that requires its object to be already filled cannot, therefore, precede the normal subject plus honorification formative on account of conflicting specifications for the 'subcat' value. Otherwise, no conflict exist.

Other kinds of restrictions on the possible word order of Japanese honorific predicate constituents can likewise be explained in the proposed framework.

6. Anaphora Resolution in Honorific Contexts

In Japanese honorific contexts, many human anaphors can be resolved by recourse to pragmatic constraints on the use of honorifics. This is an attempt to apply DR theory to the anaphora resolution in Japanse honorific contexts.

Discourse information is represented by a feature structure consisting of a set of reference markers (Universe) and a set of conditions, as in the standard version of DR (Discourse Representation) theory[7]. Fig 8.a is the initially posited DRS (Discourse Representation Structure). Addition of other discourse information to the initial DRS does not affect the theory.

```
[[univ {[[rm *speaker*[[type 'individual]]]
        [[rm *heaer*[[type 'individual]]]
        [[rm *now*[[type 'temporal-location]]]
        [[rm *here*[[type 'spatial-location]]]}]
 [conds {}]]
```

Fig 8.a. Initial DRS

[N.B.1]   Reference markers for the indexicals are directly anchored to objects in the world, but the anchoring information is not shown here.

Now let (3a) represent a discourse-initial utterance.

(3) a. *Izen ACL-88 ga hiraka-re ta toki, watasi wa aru chomei-na keisan-gengogaku-sha ni o-ai si masi ta.*
        "Once when ACL-88 was held, I met (object-honorific and performative-honorific) a certain famous computational linguist."

From this, Fig 8.b is unified as its semantic/pragmatic information. The method of specifying necessary lexical information was briefly explained in section 3.

The initial discouse information is updated by the semantic/pragmatic information of a new utterance as follows: First, DICR 1, shown in Fig 9.a below, is applied to the semantic value of a new utterance. DICR 2 is then applied to the pragmatic value. Meanwhile, anaphoric expressions in a new utterance are resolved so that the NFC[8] shown in Fig 9.b below is observed.

In this case, Fig 8.c is obtained as an updated DRS, because the type of sem|cont value is a *'basic-circumstance'* and every

```
[[sem [[cont ?x01[[reln 'meet]
                  [agent *speaker*]
                  [object ?x02]
                  [tloc ?x03]]]
   [inds {
      ?x04[[var ?x02[[type 'ind]]]
           [familiarity '-]
           [restrs {?x05[[reln 'computational-
linguist]
                     [instance ?x02]]
            ?x06[[reln 'famous]
                 [instance ?x02]]}]]]],
      ?x07[[var ?x03[[type 'tloc]]]
           [familiarity '-]
           [restrs {?x08[[reln 'hold]
                         [object ?x09]
                         [tloc ?x03]]
            ?x10[[reln 'temporally-precedes]
                 [ante ?x03]
                 [post *now*]]}]]]]
      ?x11[[var ?x09[[type 'ind]]]
           [familiarity '-]
           [restrs {?x12[[reln 'naming]
                         [name 'acl-88]
                         [named ?x09]]}]]]]]]]
   [prag [[restrs {<?x13[[reln 'honor-up]
                        [agent *speaker*]
                        [object ?x02]],
            ?x14[[reln 'honor-up]
                 [agent *speaker*]
                 [object *hearer*]]}]]]]]]]
```

Fig 8.b. Resulting Semantic Information for (3a)

Let k be a current DRS, σ be a linguistic structure for an input utterance unified from lexical specifications, and k' be a DRS to be obtained.

DICR 1.   (i)   if σ|sem|cont is typed as a 'non-quantified-circumstance', then

k'|univ = k|univ U σ|sem|inds|var, and

k'|conds = k|conds U σ|sem|cont U σ|sem|inds|restrs.

(ii)   if σ|sem|cont is typed as a 'universally-quantified-circumstance', then

k'|univ = k|univ, and

k|conds = k|conds U {[[reln '→][ante k1][post k2]]}

where k1 and k2 are newly introduced DRSs whose information contents are specified based on the σ|sem|cont|quant|ind value and the σ|sem|cont|scope value as follows

...

...

DICR 2.   k'|univ = k|univ, and

k'|conds = k|conds U σ|prag|restrs

Fig 9.a. Discourse Information Change Rules (part)

For σ to be felicitous w.r.t. k, it is required for every index i in σ that:

(i)   if i|familiarity = '-, then i|variable ∉ k|universe.

(ii)   if i|familiarity = '+, then

(a) i|variable ∈ k|universe, and

(b) i|restriction is unifiable with k|condition.

Fig 9.b. Novelty Familiarity Condition

index in the sem|cont|inds value has a [familiarity'-] attribute in Fig 8.b.

```
([[univ {[[rm *speaker*]] [[rm *hearer*]]
         [[rm *now*]] [[rm *here*]] [[rm ?x02]]
         [[rm ?x03]] [[rm ?x09]]}]
 [conds {?x01 ?x05 ?x06 ?x08 ?x10 ?x12 ?x13 ?x14}]]]
```

Fig 9.b. Updated DRS

145

In this context, assume (3b) is uttered. Fig 8.c is its unified sem/prag values.

(3) b. *?Sono keisan-gengogaku-sha wa watasi ni aisatu si yagari masi ta.*
      "That computational linguist greeted (subject-minus-honorific and performative-honorific) me."

```
[[sem [[cont ?x15[[reln 'greet]
                  [agent ?x16]
                  [recipient *speaker*]
                  [tloc ?x17 ]]]
       [inds (?x18[[var ?x16[[type 'ind]]]
                  [familiarity '+]
                  [restrs {
                      ?x19[[reln 'computational-
                                       linguist]
                          [instance ?x16]]}]]
                ?20[[var ?17[[type 'tloc]]
                    [restrs {
                        ?21[[reln 'temporally-
                                        precedes]
                            [ante ?17]
                            [post *now*]]]]}]]]
 [prag [[restrs {?22[[reln 'honor-down]
                    [agent *speaker*]
                    [object <16>]]
                ?23[[reln 'honor-up]
                    [agent *speaker*]
                    [object *hearer*]]}]]]]]]
```
Fig 8.c. Resulting Semantic Information for (3b)

Because the index ?x18 for "sono keisan-gengogaku-sha" (that computational linguist) has a [familiarity '+] attribute based on the lexical specification for 'sono', an attempt is made to resolve it by unifying ?x16 with an element of the k|univ value, requiring that their restrictions can also be unified. It stands to reason that it can be resolved because ?x16 and ?x02 are, semantically speaking, unifiable, because their semantic restrictions are {[[reln 'computational-linguist][instance ?x16]]} and {[[reln 'computational-linguist][instance ?x02]] [[reln 'famous][instance ?x02]]} respectively, and their variable types are both 'individual', which causes no incompatibility. However, their pragmatic restrictions ({[[reln 'honor-down][agent *speaker*][object ?x16]] [[reln 'honor-up][agent *speaker*][object *hearer*]]}, and {[[reln 'honor-up][agent *speaker*][object ?x02]] [[reln 'honor-up][agent *speaker*][object *hearer*]]}) prevent ?x16 from being unified with ?x02, due to the stipulation '[[reln 'honor-up][agent ?a][object ?b]] ∧ [[reln 'honor-down][agent ?a][object ?b]] = bottom'. This anaphoric resolution therefore fails. Other ways of resolving this anaphoric expression also fail because of the incompatibility of their variable types or semantic features. In any case, utterance (3b) turns out to be infelicitous by NFC.

Unlike (3b), utterance (3b'), whose sem/prag values are the same as Fig 8.c except for [[reln 'honor-up][agent *speaker*][object ?x16]] instead of [[reln 'honor-down][agent *speaker*][object ?x16]], can be given a felicitous reading, because anaphora resolution is possible without violating NFC in this case.

(3) b'. *Sono keisan-gengogaku-sha wa watasi ni aisatu nasai masi ta.*
      "That computational linguist greeted (subject-honorific and performative-honorific) me."

[N.B.] Our DICRs with NFC also explain the failure of coindexing "sono keisan-gengogaku-sha" in (4b) with a universally quantified expression "dono ... mo" (every ...) in a previous utterance, because the reference markers introduced for a universally quantified expression are in subordinate DRSs by DICR 1 and not accessible from "sono keisan-gengogaku-sha" as a possible antecedent.)

(4) a. *Izen ACL-88 ni sanka si ta toki, watasi wa* <u>dono</u> *chomei-na keisan-gengogaku-sha ni* <u>mo</u> *o-ai si masi ta.*
      "When I once took part in ACL-88, I met (object-honorific and performative-honorific) every famous computational linguist. "

    b. ? *Sono keisan-gengogaku-sha wa watasi ni aisatu nasai masi ta.(3b')*

Though many issues rermain unaddressed concerning anaphora resolution in Japanese honorific contexts, these can be approached by use of the proposed model. This model regards discourse understanding as the process of unifying various kinds of partial information, including contextual information.

7. Conclusion

A unification-based approach to Japanese honorifics based on a version of HPSG was proposed. Utterance parsing is based on the lexical specifications of a range of honorifics using a parser capable of unifying cyclic feature structures. The developed parser constitutes an important part of NADINE (NAtural Dialogue INterpretation Expert), an experimental system which translates Japanese-English telephone and inter-keyboard dialogues.

References

[1] Pollard, Carl & Ivan Sag, 1987, Information-Based Syntax and Semantics. vol. 1. CSLI Lecture Notes 13.
[2] Gunji, Takao. 1987. Japanese Phrase Structure Grammar. Reidel.
[3] Mizutani, Sizuo., 1983, "Taiguu Hyougen no Sikumi." (Structure of Honorific Expressions), in Unyou (The Pragmatics). Asakura.
[4] Harada, S. I., 1976, "Honorifics." in Shibatani (ed.), Syntax and Semantics 5. Academic Press.
[5] Wroblewski, David A., 198, "Nondestructive graph unification." in the sixth conf. on AI.
[6] Kogure, Kiyusi, et al. 1988 (forthcoming), "A Method of Analyzing Japanese Speech Act Types." in the 2nd conf. on Theoretical and Methodological Issues in Machine Translation of Natural Languages.
[7] Kamp, Hans., 1981, "A Theory of Truth and Semantic Representation." in Groenendijk et al. (eds.), Formal Methods in the Study of Language. Mathematisch Centrum.
[8] Heim, Irene. 1983, "File Change Semantics and the Familiarity Theory of Definiteness." in Bäuerle et al. (eds.), Meaning, Use and Interpretation of Language. Walter de Gruyter.