Computational Linguistics and Chinese Language Processing
Vol. 11, No. 4, December 2006, pp. 333-348

Multiply Quantified Internally Headed Relative Clause in Japanese: A Skolem Term Based Approach

Rui Otake*, and Kei Yoshimoto*

Abstract

This paper presents an analysis of Internally Headed Relative Clause (IHRC) construction in Japanese within the framework of Combinatory Categorial Grammar [Steedman 2000]. Shimoyama [1999] argues that when an IHRC appears within the scope of a universal quantifier, the interpretation of the IHRC exemplifies E-type anaphora and that the LF representation of the IHRC should have a variable bound by the quantifier in the matrix clause. To accommodate this argument Shimoyama posits a free variable of a functional type to which the bound variable is applied, and whose denotation is determined by the context-dependent assignment function. However, since there is in principle no limit to the number of quantifiers in the matrix clause (and accordingly that of bound variables in the IHRC), the semantic type of the free variable would be highly ambiguous if the IHRC occurs within the scope of multiple quantifiers. The current analysis assumes that the interpretation of IHRCs exhibits an instance of generalized Skolem term [Steedman 2005], a term whose denotation varies with the value of bound variables introduced by scope-taking operators, but which is interpreted as a constant in the absence of such operators. This paper provides a straightforward account for the semantics of the construction without invoking the complexities of the type ambiguity of free variables.

Keywords: Combinatory Categorial Grammar, Generalized Skolem Term, Internally Headed Relative Clause, Japanese, Quantification

1. Introduction

This paper presents an analysis of Internally Headed Relative Clause (IHRC) construction in Japanese paying particular attention to the effect of quantification on its interpretation. (1)

* Graduate School of International Cultural Studies, Tohoku University, 41 Kawauchi, Aoba-ku, Sendai 980-8576, Japan. Tel: +81-22-795-7550, Fax: +81-22-795-7850
E-Mail: otake@linguist.jp; kei@insc.tohoku.ac.jp

[Received April 30, 2006; Revised October 19, 2006; Accepted November 7, 2006]

illustrates the basic form of the construction:[1]

Taroo-ga [Hanako-ga ringo-o muita] no-o tabeta.
Taro-NOM Hanako-NOM apple-ACC peeled NML-ACC ate (1)
'Taro ate the apple that Hanako peeled.'

The bracketed clause *Hanako-ga ringo-o muita* 'Hanako peeled an apple' is followed by the nominalizer *no* and the accusative particle *-o*, thereby construed as the object of the matrix verb *tabeta* 'ate'. Since the verb requires as its semantic restriction that the accusative argument be an edible thing, it anaphorically picks up the referent of *ringo* 'apple' from the embedded clause. This kind of construction is often contrasted with the Externally Headed Relative Clauses (EHRC), which is illustrated in (2).

Taroo-ga [Hanako-ga muita] ringo-o tabeta.
Taro-NOM Hanako-NOM peeled apple-ACC ate (2)
'Taro ate the apple that Hanako peeled.'

As can be seen from the translation, (1) and (2) have almost the same meaning. But we hasten to add that IHRCs are not always paraphrasable to the EHRC version, for the former construction is subject to some pragmatic condition for its felicitous use, which we will not attempt to specify. A terminological note: we use the term the *antecedent* of an IHRC to mean the referent of the IHRC which functions as the argument of the matrix predicate. And we also use the term the *head* of the IHRC to refer to the linguistic element in the IHRC which describes the antecedent. For example, the antecedent of the IHRC in (1) is the apple that Hanako peeled, and the head is the noun *ringo* 'apple'. It is important to notice that we define IHRC construction in terms of the nominal character of its anaphoric referent, despite the name suggesting that the presence of the head noun inside the relative clause is the defining feature of the construction. In fact, there are cases where the IHRC has no explicit nominal head. Such examples will be dealt with in section 2.2.

This paper is organized as follows. In section 2, we review the observation made by Shimoyama [1999] and her E-type analysis of IHRC. We then address the problem that the E-type analysis would raise focusing on multiply quantified IHRCs. We also discuss the

[1] Abbreviations used: ACC = accusative, ALL = allative, CL = classifier, COMP = complementizer, COP = copula, GEN = genitive, LOC = locative, NML = nominalizer, NOM = nominative, TOPIC = topic.

interpretational characteristics of IHRC, drawing on the study of Kikuta [2000]. Section 3 introduces the notion of generalized Skolem term, which provides a straightforward account for multiply quantified IHRC. Finally, section 4 concludes.

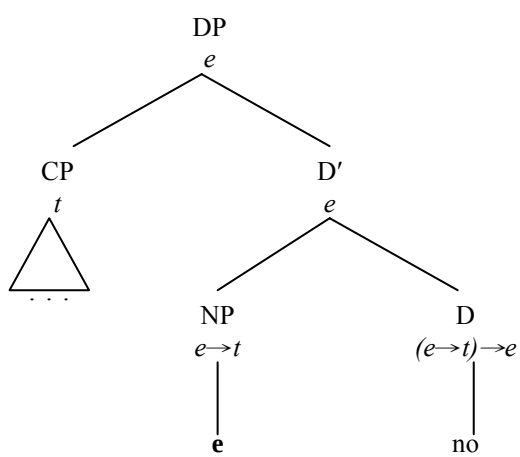
2. Previous Analysis and Its Problem

2.1 Shimoyama's [1999] E-type analysis

Shimoyama [1999] claims that when IHRC appears within the scope of a universal quantifier, the interpretation of IHRC exemplifies E-type anaphora:

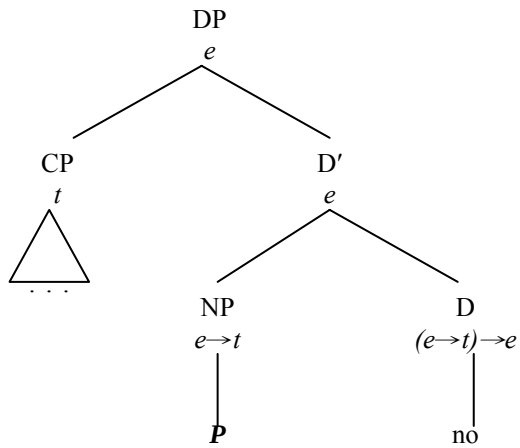
Dono gakusei$_i$-mo [soitu$_i$-ga kongakki peepaa-o san-bon kaita] no-o
Every student he-NOM this:semester paper-ACC three-CL wrote NML-ACC
kesa teisyutusita. (3)
this:morning turned:in
'This morning every student turned in the three term papers he or she wrote this semester.'

In (3), the subject of the embedded clause is bound by the universal quantifier in the matrix clause. Note that "[t]he matrix object [...] does not refer to any particular set of term papers [Shimoyama 1999]," and the interpretation of the matrix object can be paraphrased by the definite description *the term papers he or she wrote this semester*. The interpretation of the relative clause is an instance of E-type anaphora [Evans 1980]. Given this observation, Shimoyama [1999] proposes the LF structure of IHRC as schematized in (4). According to this analysis, CP in the Spec of DP corresponds to the relative clause, which moves up to some higher position and is interpreted independently because of type mismatch. The DP is headed by the nominalizer *no*. The complement NP indicated by e is an empty pronoun. Basically, the interpretation of this empty pronoun determines the argument of matrix predicate.



(4)

For example, in the LF representation (5) of the IHRC in (1), the empty pronoun is represented as *P*, a free variable of type *e*→*t*.

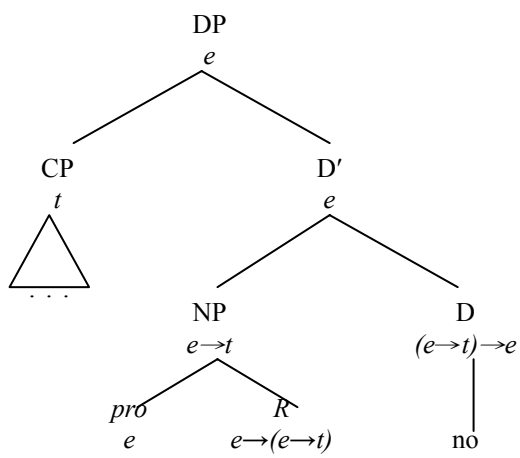


(5)

Now, it is assumed that free variables in general are assigned a value by the assignment function *g* relative to the context *c*. In this case, P is assigned the value in (6): the set of apples which Hanako peeled.

$$[\![P]\!]_{g_c} = \lambda x.apple'x \wedge peel'x\ hanako' \quad (6)$$

Shimoyama assumes that the nominalizer *no* is interpreted as the function from a set to the maximality of the set, adopting Link's [1983] analysis of definite descriptions.[2] As a



(7)

[2] Note in passing that our analysis in section 3 assumes that plurals are translated as set individuals.

result, the denotation of IHRC is equivalent to the English definite description *the apples that Hanako peeled*. In the case of (3), the LF representation of the IHRC is something like (7), in which the empty pronoun is further divided into *pro* and a free variable *R*.

pro is a variable of type *e* bound by the universal quantifier in the matrix clause. On the other hand, *R* is a free variable of type $e\rightarrow(e\rightarrow t)$, to which an assignment function g_c assigns the value in (8) as a salient two-place predicate in the context.

$$[\![R]\!]_{g_c} = \lambda x.\lambda y.paper'y \wedge wrote'yx \tag{8}$$

The latter takes *pro* as its argument to yield the interpretation shown in (9).

$$[\![R]\!]_{g_c}\left([\![pro]\!]_{g_c}\right) = \lambda y.paper'y \wedge wrote'yx \tag{9}$$

In words, this is the set of *y* such that *y* is a set of papers that *x* wrote, where *x* is bound by the universal quantifier. Then the result serves as the argument for *no* as before. In the end, (3) is interpreted as 'for every student *x*, *x* wrote three papers this semester, and this morning *x* turned in the papers *x* wrote this semester.' We thus get correct semantics. However, this analysis poses two problems. First, as is noted by Shimoyama herself, the assignment function is not properly constrained. It just picks up a property or relation which is 'salient' in the context. We will discuss this problem in section 2.2. We will assume that the appropriate constraints can be captured by the predicate *result'* or *abt'* although we will not be determinate on how to decide between these two options. The second problem is the ambiguity in the semantic type of the free variable. As we have just seen, the assignment function g_c assigns a value to a free variable relative to the context. In order to assign a value, at least the semantic type of the variable needs to be known. However, the semantic type of that free variable can be determined only after the context is available. Then, the problem is that it is unclear as to how the context can be available before the context-dependent interpretation comes in. Furthermore, the semantic type of the free variable can be arbitrarily complex according to the number of universal quantifiers in the matrix clause. This is illustrated in (10-11):

Dono sensei-mo subeteno zyugyoo-de [menomae-de gakusei-ga neteiru]
every prof all class-LOC before:eyes-LOC student-NOM sleeping
no-o tatakiokosita (10)
NML-ACC woke:up:roughly
'Every prof, in all his classes, woke up a student who is sleeping before his eyes.'

Dono bando-no dono gitarisuto-mo subeteno suteezi-de dono kyoku-demo
Every band-GEN every guitarist all stages-LOC every song-LOC
[gitaa-no tyuuningu-ga kurutteiru] no-o sonobade tatakikowasita (11)
guitar-GEN tuning-NOM wrong NML-ACC on:the:spot broke
'Every guitarist of every band smashed the guitar which was out of tune in every song on all stages.'

In principle, there is no limit to the number of quantifiers in the matrix clause. Therefore, a mechanism that does not invoke complexity of this sort would be preferred. In section 3, we propose an analysis that can derive the interpretation of such multiply quantified IHRC in the same way as the quantifier-free cases like (1) and singly quantified cases like (2) by introducing the concept of the generalized Skolem term proposed in Steedman [2005].

2.2 The Interpretational Characteristics of IHRC

In order to state a proper constraint on the possible antecedent of an IHRC, let us examine the semantics of IHRC in more detail. The anaphoric nature of IHRC is best illustrated by the fact that the antecedent of an IHRC is occasionally not expressed as a linguistic element. (12-14) are such examples of 'headless' IHRC from Nomura [2000].[3]

[Nikai-de suisoo-ga ahureta] no-ga sita-ni
second:floor-LOC fish:tank-NOM overflowed NML-NOM downstairs-ALL
morete hita (12)
leak come
'The fish tank upstairs overflowed and (the water) leaked to downstairs.'

[Kesa kao-o sotta] no-ga yuugata-niwa mata nobite kita
this:morning face-ACC ahaved NML-NOM evening-TOP again growing came (13)
'I shaved my face in the morning, and (the beard) started to grow again in the evening.'

[Tuti-o hotta] no-o ue-kara nozokikonda
soil-ACC dug NML-NOM up-from looked:into (14)
'I dug the soil, and looked into (the hole).'

[3] In the translation, antecedents are given in brackets.

Determining the antecedent of a headless IHRC obviously requires inference of some sort. Kikuta [2000] addresses the question of exactly how much information is needed for this inference within the framework of Generative Lexicon (GL). To recapitulate Kikuta's argument, if the inference always requires unconstrained pragmatics, the theory would overgenerate IHRCs with an illicit antecedent. Accordingly, there should be a more restricted way for determining the antecedent. And indeed she shows that the antecedent can be identified only by the linguistically specified information. Kikuta's conclusion is that the possible antecedent of the headless IHRC must meet the following conditions: (i) necessary involvement in the event described by the main predicate in the IHRC; (ii) recoverability of the denotation from linguistically pre-specified information; and (iii) presence in the resultant state of the process described by the main predicate in the IHRC. Note that for example (12), the antecedent is the water which is necessarily involved in the overflowing event, lexically specified by the predicate *ahureta* 'overflowed', and exists in the resultant state. Other examples can also be shown to satisfy these conditions. In contrast, violating these conditions leads to unacceptability. Following Kikuta's result (but simplifying the matter somewhat), in our analysis we will use a predicate *result′* which is defined such that *result′ px* means that an element *x* is involved in the resultant state of the process or event described by the proposition *p*. However, the three conditions given above cannot straightforwardly be applied to the IHRCs in general. Consider (15), the example taken from a news article (asahi.com, June 22, 2004):

[Denwa-o kawatta betuno otoko-ga nakizyakutteita] no-o syuhu-wa
phone-ACC got the:other man-NOM sobbing NML-ACC housewife-TOP
kaisyain-no otto da to omoikonda. (15)
office:worker-GEN husband COP COMP believed
'The housewife took the other man sobbing on the phone as her husband, an office worker.'

The main predicate of the IHRC is *nakizyakutteita* 'sobbing', which denotes an activity, apparently lacks a lexical specification of the resultant (or consequent) state. Therefore, there is no way to satisfy condition (iii). And yet the antecedent *betuno otoko* 'the other man' is an entity necessarily involved in the described event, in accordance with the above condition (i). In such cases, we will use a predicate *abt′* for 'aboutness' relation defined informally such that *abt′ px* means that an element *x* is necessarily involved in the event described by the proposition *p*. Now we have two distinct relations (*i.e. result′* and *abt′*) to describe the semantic constraint on the possible antecedent of IHRCs. Note that *result′* is a subtype of *abt′*

since the former entails the latter. We will define the nominalizer category as initially having the interpretation containing the predicate *abt′*, which is on occasion replaced with *result′*. In the following, we give an informal sketch of how this replacement is done. However, in the analysis in section 3, we will take this replacement for granted, treating *result′* as if it is given lexically. Given the fact that an IHRC is syntactically a complete sentence, we can assume that an IHRC constitutes a separate information unit from the matrix clause. And we adopt the claim of Segmented Discourse Representation Theory [Asher and Lascarides 2003] that every information unit or proposition is connected via some rhetorical relation in order for the entire discourse to be coherent. Then the IHRC and the matrix clause must be connected via some rhetorical relation. By way of illustration, let us consider (12) again. Here, we have two clauses, namely the IHRC and the matrix clause, and the two corresponding propositions: the fish tank upstairs overflowed (π_1), and *something* leaked to downstairs (π_2). In the latter proposition, *something* corresponds to the nominative argument IHRC. Being anaphoric, this must be resolved in some way. In addition, the two propositions need to be connected via some rhetorical relation. Suppose that the latter requirement is somehow fulfilled by inferring *Result*(π_1,π_2) as the relevant rhetorical relation. Then the semantics of *Result* entails that the event of π_1 caused that of π_2. Now, *something* in π_2 can plausibly be equated with the water, making the discourse (consisting of two clauses) coherent. We assume that the relation symbol *abt′* originating from the lexical information is replaced with the more specific relation symbol *result′* in the process of such inference. Note, incidentally, that this kind of approach may provide a way to account for the Relevancy Condition on IHRC discussed in Kuroda [1975-6]:

> For a p.-i. relative clause [IHRC] to be acceptable, it is necessary that it be interpreted pragmatically in such a way as to be directly relevant to the pragmatic content of its matrix clause.

As the effect of the Relevancy Condition, events described by an IHRC and the matrix clause are typically related in any of the following terms: (i) temporal overlap, (ii) relevance of purpose, (iii) relevance of motivation, or (iv) spatial proximity. Note that Kuroda stated the Relevancy Condition as a mere descriptive generalization, whereas in our approach sketched above, this can be viewed as a consequence of the principle of discourse coherence in general, since if IHRC cannot be connected by some rhetorical relation, there would be no way to attach the information of IHRC to the matrix clause in a coherent manner. The effects of the Relevancy Condition mentioned above can also be regarded as entailments of the inferred rhetorical relation.

3. Generalized Skolem Term Analysis

3.1 Generalized Skolem Term

In this section, we will introduce the concept of the generalized Skolem term proposed by Steedman [2005]. We briefly sketch the basic idea here and we will show in section 3.3 how to apply it to the analysis of (multiply quantified) IHRCs. The basic idea is this: a generalized Skolem term is a term whose denotation varies with the value of bound variables introduced by scope-taking operators such as universal quantifiers, but which is interpreted as a constant in the absence of such operators. One of the main motivations for such a mechanism is to give an analysis of the alternation of quantifier scope. By way of illustration, let us consider the sentence *Everybody loves somebody*. The narrow reading of *somebody* can be translated as

$$\forall x\left[\, person'x \rightarrow \exists y[\, person'y \wedge love'yx]\right] \tag{16}$$

However, we can entirely eliminate the existential quantifier from Logical Form, by replacing the existentially quantified variables with the Skolem term *sk′x*.

$$\forall x\left[\, person'x \rightarrow \left(\, person'(sk'x) \wedge love'(sk'x)x\right)\right] \tag{17}$$

Note that *sk′* here is a Skolem function, and the referent of *sk′x* is dependent on who the variable *x* refers to. On the other hand, we get the wide scope reading of *somebody* by letting the Skolem term be a constant *sk′*.

$$\forall x\left[\, person'x \rightarrow \left(\, person'sk' \wedge love'sk'x\right)\right] \tag{18}$$

The above transformations of Logical Form illustrate the standard Skolemization. In the current framework, however, indefinite noun phrases are interpreted as Skolem terms right from the start, and the nominal properties such as *person′* will be directly associated with them. More specifically, when introduced as a Logical Form element, a Skolem term is indicated as *skolem′p* where *p* designates a nominal property. At this stage, this is *unspecified* as to its arguments. At some later step in the derivation, an operation called *Skolem specification* is applied to this unspecified Skolem term to yield a *generalized Skolem term*, designated as sk_p^E, where *E* is the *environment*, an ordered set consisting of the bound variables of the universal quantifiers that take scope at that point of the derivation. Since sk_p is a function with *E* as its argument, its reference varies depending on the values of bound variables. And if *E* is the empty set, the sk_p^E will be a constant. The notion of environment is incorporated in the grammatical rule of Combinatory Categorial Grammar (CCG) in the following way. First, we have two function application rules:[4]

---

[4] In this paper, we use only the application rules, which (apart from the notational convention) is common to all variants of Categorial Grammars. The reason for this choice is just the simplicity of presentation. For other rules of CCG, we refer the reader to Steedman [2000].

$$X/Y : f' \quad Y : a' \quad \Rightarrow \quad X : f'a' \qquad (>)$$

$$Y : a' \quad X\backslash Y : f' \quad \Rightarrow \quad X : f'a' \qquad (<)$$

The environment is the operator bound variable identifier and is associated with the propositional body of the interpretation. For example, the transitive verb *loves* has the following category with its environment being the empty set:

$$\text{loves} := (S \backslash NP_{3S}) / NP : \lambda x.\lambda y.[loves'xy]^{\{\}} \qquad (19)$$

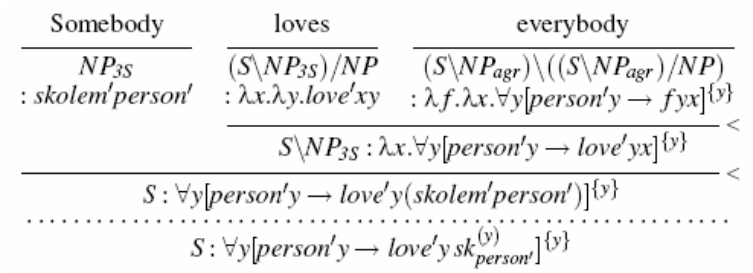
Application of the rule induces environment passing in the following way: if a function with environment F is applied to an argument with environment A, the environment of the argument in the resulting Logical Form is the union of the two ($F \cup A$). We often omit the environment from the notation where it is of little interest. Let us now turn to the way this works. Expressions traditionally analyzed as an existential quantifier such as *somebody* are analyzed here as an unspecified Skolem Term *skolem'person'*. In the sentence *Everybody loves somebody*, for example, if specification applies at a point of derivation in which the Skolem term has not yet been in the scope of the universal quantifier as in (20), the resulting generalized Skolem term will be $sk_{person'}$, hence we get the wide scope reading of *somebody*. In the derivation shown below, Skolem specification is indicated by the dotted underline.

$$\dfrac{\dfrac{\text{Everybody}}{S/(S\backslash NP_{3S}) : \lambda f.\forall y[person'y \rightarrow fy]^{\{y\}}} \quad \dfrac{\dfrac{\text{loves}}{(S\backslash NP_{3S})/NP : \lambda x.\lambda y.love'xy} \quad \dfrac{\dfrac{\text{somebody}}{NP : skolem'person'}}{NP : sk_{person'}}}{S\backslash NP_{3S} : \lambda y.love'sk_{person'}y}>}{S : \forall y[person'y \rightarrow love'sk_{person'}y]^{\{y\}}}> \qquad (20)$$

In contrast, if specification applies after it enters within the scope of the universal quantifier, the resulting generalized Skolem term is $sk^{(y)}_{person'}$, where $sk_{person'}$ is a function which takes the variable y bound by the universal quantifier as its argument, hence the narrow scope reading of somebody. The following illustrates the derivation of this reading.

$$\dfrac{\dfrac{\dfrac{\text{Everybody}}{S/(S\backslash NP_{3S}) : \lambda f.\forall y[person'y \rightarrow fy]^{\{y\}}} \quad \dfrac{\dfrac{\text{loves}}{(S\backslash NP_{3S})/NP : \lambda x.\lambda y.love'xy} \quad \dfrac{\text{somebody}}{NP : skolem'person'}}{S\backslash NP_{3S} : \lambda y.love'(skolem'person')y}>}{S : \forall y[person'y \rightarrow love'(skolem'person')y]^{\{y\}}}>}{S : \forall y[person'y \rightarrow love'sk^{(y)}_{person'}y]^{\{y\}}} \qquad (21)$$

The scope alternation is also observed if the universal quantifier noun phrase appears in the object position. This is accounted for in a quite similar way to that shown above. The derivation below illustrates the inverse scope reading.



(22)

3.2 Distributivizing Verb Category in Japanese

Before going into the Skolem term approach to IHRC, let us consider how the scope alternation is achieved in Japanese. One striking fact about scope alternation in Japanese is that its availability is more restricted than in English. While in English the universal quantifier in the object position can take the inverse scope over the subject indefinite as in *Somebody loves everybody*, the Japanese counterpart does not seem to accept such reading. In (23), we observe that the Skolem term *dareka* 'somebody' allows only the narrow scope reading (*cf.* Nakamura 1993):

Dareka-ga daremo-o aisiteiru.
Somebody-NOM everybody-ACC love (23)
'Somebody loves everybody'

However, if the object NP is scrambled to the sentence initial position, and precedes the subject indefinite *dareka* 'somebody', both narrow and wide scope readings are available.

Daremo-o dareka-ga aisiteiru.
Everybody-ACC somebody-NOM love (24)
'Somebody loves everybody'

This situation is quite unlike English examples. To accommodate this fact, we tentatively adopt Steedman's suggestion (2005:54, fn.51) that languages such as Japanese entirely lack generalized quantifier NPs and that the work of the universal quantifier in such languages are

done by distributivizing verb categories. We assume that *daremo* 'everybody' denotes a set of individuals whose members are all the people in the universe of discourse, indicated as *all-people'*. (25) illustrates the derivation of (24), in which *dareka* 'somebody' is interpreted as having narrow scope.

$$\dfrac{\dfrac{\dfrac{\text{Daremo-o}}{NP_{ACC} : all\text{-}people'} \quad \dfrac{\text{dareka-ga}}{NP_{NOM} : skolem'person'} \quad \dfrac{\text{aisiteiru}}{(S_{DIST}\backslash NP_{ACC})\backslash NP_{NOM} : \lambda y.\lambda x.\forall w[w \in x \rightarrow love'wy]^{\{w\}}}}{S_{DIST}\backslash NP_{ACC} : \lambda x.\forall w[w \in x \rightarrow love'w(skolem'person')]^{\{w\}}}<}{S_{DIST} : \forall w[w \in all\text{-}people' \rightarrow love'w(skolem'person')]^{\{w\}}}< \quad (25)$$
$$\vdots$$
$$S_{DIST} : \forall w[w \in all\text{-}people' \rightarrow love'w\, sk^{(w)}_{person'}]^{\{w\}}$$

Here, we defined the category of the verb so that each member of the set *all-people'* distributes over the Skolem term introduced by the subject NP. If we further assume that distributivizing as well as scrambling is realized by a lexical rule, then the fact that distributive reading is absent in a non-scrambled sentence can be thought of an accidental lack of such lexical rule. The derivation (26) illustrates the only possible reading for (23).

$$\dfrac{\dfrac{\dfrac{\text{Dareka-ga}}{NP_{NOM} : skolem'person'}}{NP_{NOM} : sk_{person'}} \quad \dfrac{\dfrac{\text{daremo-o}}{NP_{ACC} : all\text{-}people'} \quad \dfrac{\text{aisiteiru}}{(S\backslash NP_{NOM})\backslash NP_{ACC} : \lambda x.\lambda y.love'xy}}{S\backslash NP_{NOM} : \lambda y.love'all\text{-}people'y}<}{S : love'all\text{-}people'sk_{person'}}< \quad (26)$$

3.3 IHRC as Generalized Skolem Term

The current analysis views the interpretation of IHRC as an instance of generalized Skolem term, and provides a straightforward account for the semantics of the construction. In order to capture the restriction on the interpretation of IHRC, the nominalizer category is defined as in (27), where *R* stands for either *abt'* or *result'* (see section 2.2 for the discussion of this treatment):

$$\text{no} := NP\backslash S : \lambda p.skolem'(\lambda x.Rpx) \quad (27)$$

The derivation of (1) is shown in (28).

$$\dfrac{\dfrac{\dfrac{\text{Taroo-ga}}{NP : taro'} \quad \dfrac{\dfrac{\text{Hanako-ga ringo-o muita}}{S : peel'apple'hanako'} \quad \dfrac{\text{no-o}}{NP\backslash S : \lambda p.skolem'\lambda x.result'px}}{NP : skolem'(\lambda x.result'(peel'apple'hanako')x)}< \quad \dfrac{\text{tabeta}}{(S\backslash NP)\backslash NP : \lambda y.\lambda z.eat'yz}}{S\backslash NP : \lambda z.eat'skolem'(\lambda x.result'(peel'apple'hanako')x)z}<}{S : eat'skolem'(\lambda x.result'(peel'apple'hanako')x)taro'}< \quad (28)$$
$$\vdots$$
$$S : eat'sk_{\lambda x.result'(peel'apple'hanako')x}taro'$$

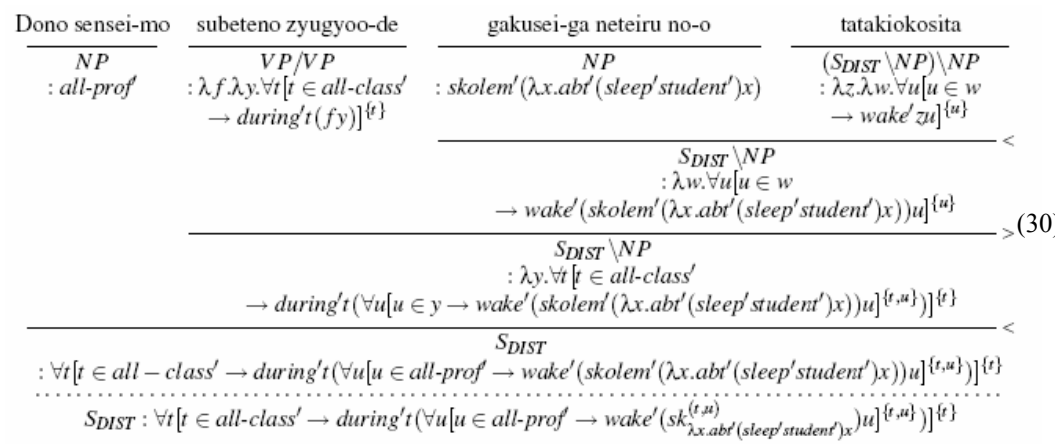
The semantics of the embedded clause is fairly obvious. This is then applied to the function defined in (27). The 'result' predicate *result′ px* indicates that the event described by *p* yields an individual *x*. Thus the IHRC is analyzed as a Skolem term with the property of being involved in the result state of Hanako's peeling an apple. It is further subject to Skolem specification (indicated by the dotted line as before). Since there is no universal quantifier, the resulting generalized Skolem term is a constant whose referent is the apple that Hanako peeled. It should be obvious that this analysis also accommodates to the examples of headless IHRCs discussed in section 2.2. Interpretation of an IHRC is always determined via the *abt′* or *result′* relation, regardless of whether the antecedent is explicitly expressed or not. Still, we have to admit that the nominal property of the Skolem term $\lambda x.\ result'(peel'apple'hanako')x$ is not a sufficient characterization of the antecedent, as it can also be applied to the apple skin, rather than only to the peeled apple fruit. Basically this λ-expression should function as a constraint on the possible referent of the IHRC, and it is subject to the process of anaphora resolution. The resolved Skolem term in this case would be $skolem'\lambda x.(result'(peel'apple'hanako')x \wedge fruit'x)$. However, we will gloss over this problem here, and the question of how this mechanism works is left open for future research. Now, let us examine the interpretation of the IHRC that occurs within the scope of a universal quantifier. The derivation of (3) is shown in (29):

$$\dfrac{\dfrac{\text{Dono gakusei-mo}}{\underset{: all\text{-}student'}{NP}} \quad \dfrac{\dfrac{\text{soitu-ga peepaa-o san-bon kaita no-o}}{\underset{: skolem'(\lambda x.result'(write'3papers'(pro'u))x)}{NP}} \quad \dfrac{\text{teisyutusita}}{\underset{: \lambda z.\lambda w.\forall u[u \in w \rightarrow turn\text{-}in'zu]^{\{u\}}}{(S_{DIST}\backslash NP)\backslash NP}}}{S_{DIST}\backslash NP : \lambda w.\forall u[u \in w \rightarrow turn\text{-}in'(skolem'(\lambda x.result'(write'3papers'(pro'u))x))u]^{\{u\}}}<}{S_{DIST} : \forall u[u \in all\text{-}student' \rightarrow turn\text{-}in'(skolem'(\lambda x.result'(write'3papers'(pro'u))x))u]^{\{u\}}}< \quad (29)$$

$$S_{DIST} : \forall u[u \in all\text{-}student' \rightarrow turn\text{-}in'sk^{(u)}_{\lambda x.result'(write'3papers'(pro'u))x}u]^{\{u\}}$$

Here, *soitu* in the IHRC is interpreted as *pro′u*. As the notation may suggest, it is introduced as an ordinary pronoun. This will later get bound by the universal quantifier.[5] The last line is the result of Skolem specification. Since the Skolem term is in the scope of the universal quantifier, it takes the bound variable *u* as its argument. Therefore, the resulting generalized Skolem term refers to different sets of term papers, according to the value of *u*. This is the desired result, and this is achieved without invoking any complexity such as the type ambiguity of the empty pronoun. Multiply quantified cases like (10) can be derived in a similar way:

[5] We take this process as anaphora resolution of the usual kind. This is because *soitu* could also be interpreted as deictic pronoun, in which case there is a ghostwriter (referred to by *soitu*) who wrote all the papers that every student turned in (a highly implausible situation, though).



(30)

Since Skolem term automatically takes the bound variables of the environment, the problem of type ambiguity that the E-type analysis suffered does not arise.

4. Conclusion

We developed an analysis of Internally Headed Relative Clause (IHRC) construction in Japanese within the framework of Combinatory Categorial Grammar [Steedman 2000]. In section 2, we first looked at the argument made by Shimoyama [1999] and her E-type analysis of IHRC. We addressed the problem of type ambiguity that her E-type analysis would raise focusing on multiply quantified IHRCs. We then discussed the interpretational characteristics of IHRC, drawing on Kikuta's [2000] study of headless IHRCs. We generalized her idea by partly adopting Asher and Lascarides's [2003] theory, and argued that semantics of rhetorical relation would also help to determine the antecedent. And we also suggested that Kuroda's [1975-6] Relevancy Condition on IHRC can be viewed as a consequence of the principle of discourse coherence in general. In section 3 we introduced the notion of generalized Skolem term, and discussed the problem of scope alternation in Japanese, which is quite different from English, and motivated the distributivizing verb category adopting the Steedman's [2005] suggestion. Finally we integrated the whole discussion into the analysis of IHRCs. The main point is that, if we take the interpretation of an IHRC as a generalized Skolem term, we can attain the uniformed analysis for the several kinds of IHRCs, namely, headless IHRC, simple (quantifier-free) IHRC, and (singly or multiply) quantified IHRC. Our focus here was exclusively on the IHRC, but of course the approach here also applies to other kinds of noun phrases. However, the theoretical implication of this approach in other nominal construction is not entirely clear at this point, and is left for future work.

Acknowledgements

The research reported here was supported in part by the Tohoku University 21st Century Center of Excellence (COE) Program in Humanities, Research and Strategic Center for an Integrated Approach to Language, Brain, and Cognition (http://www.lbc21.jp/). We are grateful to two anonymous reviewers for their comments.

References

Asher, N., and A. Lascarides, *Logics of Conversation*. Cambridge: Cambridge University Press, 2003.

Evans, G., "Pronouns," *Linguistic Inquiry,* 11, 1980, pp. 337-362.

Kikuta, C. U., "Qualia Structure and the Accessibility of Arguments: Japanese Internally-headed Relative Clauses with Implicit Target," *The Proceedings of the 14th Pacific Asia Conference on Language, Information and Computation*, Akira Ikeya and Masahito Kawamori (eds.), 2000, pp. 153-164.

Kuroda, S.-Y., "Pivot-Independent Relativization in Japanese II," 1975–1976, Reprinted in S.-Y. Kuroda, *Japanese Syntax and Semantics*. Dordrecht: Kluwer, 1992.

Link, G., "The Logical Analysis of Plurals and Mass Terms: A Lattice-theoretical Approach," in *Meaning, Use, and Interpretation of Language*, R. Bäuerle et al. (eds.), de Gruyter, Berlin, 1983, pp. 302-323.

Nakamura, M., "Scrambling and Scope in Japanese." In Patricia M. Clancy (ed.), *Japanese/Korean Linguistics*, vol. 2, 1993, pp. 283–298. Stanford Linguistics Association, Stanford CA: CSLI.

Nomura, M., *The Internally-Headed Relative Clause Construction in Japanese: A Cognitive Grammar Approach*. Ph.D. dissertation, University of California, San Diego, 2000.

Shimoyama, J., "Internally Headed Relative Clauses in Japanese and E-type Anaphora," *Journal of East Asian Linguistics* 8, 1999, pp. 147-182.

Steedman, M., *The Syntactic Process*. Cambridge, MA: MIT Press, 2000.

Steedman, M., *Surface-Compositional Scope-Alternation without Existential Quantifiers*. Draft, 2005. http://www.iccs.informatics.ed.ac.uk/˜steedman/papers.html