# SYNTACTIC CONSTRAINTS ON RELATIVIZATION IN JAPANESE

SIRAI, HIDETOSI

School of Computer and Cognitive Sciences
Chukyo University
101 Tokodate, Kaizu-cho, Toyota, Aichi, Japan 470-03

TOMIOKA, YUTAKA

Faculty of Engineering
University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan 113

## Abstract

This paper discusses the formalization of relative clauses in Japanese based on JPSG framework. We characterize them as *adjuncts* to nouns, and formalize them in terms of constraints among grammatical features. Furthermore, we claim that there is a constraint on the number of slash elements and show the supporting facts.

## 1 Introduction

This paper discusses the formalization of relative clauses in Japanese based on JPSG (Japanese Phrase Structure Grammar)[1, 2], which is a constraint-based grammar formalism like HPSG(Head-driven Phrase Structure Grammar)[7, 8]. We have worked on JPSG with Prof. Gunji, and also have developed a parser based on an efficient mechanism for dealing with constraints[3, 4, 10] to show that JPSG is effective even for the computational processing of natural language.

In the next section, we briefly introduce JPSG theory. Following a simple characterization of relative clauses in Japanese language in section 3, we disscuss the variety of acceptability in secton 4, and describe its formalization in terms of constraints among the grammatical features in section 5. And in section 6 we will claim that there is a constraint on the number of slash elements and show the supporting facts.

## 2 Basics of JPSG

This section describes a brief introduction to JPSG, which is a grammar formalism originally for the Japanese language. As with HPSG, JPSG is *feature based* and *constraint based* grammar.

### 2.1 Features

Grammatical categories are represented as sets of features. We list the features used in this paper.

(1) Features used in this paper

pos (part of speech) same as in HPSG.

gr (grammatical relation) takes either *subj*, *obj*, or *iobj* as the value.

subcat (subcategorization) designates the set of categories (complements) that a particular category (head) requires. Though we have to distinguish two types of complements (*ie.*, agglutinated or not) in Japanese, for simplicity, we assume that subcat designates the both types of complements.

dep (dependent) designates the category that a particular category (adjunct) modifies.

core roughly corresponds to **CAT** feature in HPSG [8]. The value is a set of features including subcat feature and so-called *head* features such as pos, gr and dep.

slash designates a syntactic gap within the grammatical category involved.

sem (semantics) designates the semantic representation of the grammatical category involved.

In this paper, categories are designated by a left square bracket ("[") followed by an indefinite number of feature specifications (a feature name followed by its value) separated by commas(",") followed by a right square bracket ("]"). When the value is null or not relevant, the entire feature specification can be omitted. The sharing structure is indicated by variables such as $X, Y, \ldots$, which is distinguished by an initial capital letter as in programming language Prolog. Finally, a category of the form [core {pos $c$, $\ldots$}, $\ldots$, sem $s$] is often abbreviated as $c[\ldots]{:}s$.

Because grammatical relations play no role in ordering complements in Japanese, we assume only a binary-branching phrase structure schema as seen in the next subsection.

### 2.2 The Phrase Structure Schema and Grammatical Principles

We are assuming three basic phrase structure schema for Japanese: *complementation*, *adjunction*, and *co-*

*ordination:*

(2) a. *Complementation*  $M \rightarrow C \quad H$
  b. *Adjunction*  $M \rightarrow A \quad H$
  c. *Coordination*  $M \rightarrow H_1 \quad H_2$

where M stands for an arbitrary mother category, C a complement, H a head category, and A an adjunct. Each category is construed as complex symbols, or *features*, with internal structures. The above structures are uniquely characterized by the features. For example, the head in the complementation structure should have **subcat** feature one of whose value is unified with the complement, C.

Furthermore, we assume the following grammartical principles, which are applied to every structures:

**Head Feature Principle:** same as in HPSG. We assume that **pos, dep,** and other several features are head features.

**Subcat Feature Principle:** In the complementation structure shcema (2b), the value of **subcat** of the mother category unifies with that obtained by subtracting the left daughter category from the value of **subcat** of the head category. In the other strucrues, the value of **subcat** of the mother category unifies with that one of the head category.

**Binding Feature Principle:** The value of a binding feature of the mother category unifies with the union of its value of the left daughter category and its value of the head category minus the category bound of this local branching[1]. We assume that **slash** and some other features are binding features.

**Slash** feature and Binding Feature Principle play a crucial role in describing so-called *long distance dependency* phenomena. We will discuss them in detail in the following sections.

## 3 Relative Clauses in Japanese

Traditionally the relative clause in Japanese has been considered as a kind of so-called *ren-tai syusyoku setu* 'the clause for modifying noun'. Because it is characterized as an embedded sentence for modifying the head noun in the structure $[_S \ldots]N$, it can be considered as an adjunct to the head noun.

In this paper, we adopt this idea. Thus, the structure of relative clause–head noun construction can be described as shown in Figure 1, where the left daughter category is a relative clause, and the right daughter category is a head noun. The fact that the left daughter category has **dep** feature whose value

---

[1]For example, when the left daughter and the head daughter have { A } and { B } as their slash feature value respectively, the value of slash of the mother will be { A, B } in most cases. However, in case that A can be unifiable with B, the result can be { A }. Furthermore, in case that A (or B) is bound in the local branching structure (as in Figure 4), the result will be { B } (or { A }).
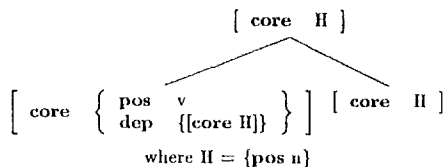


where $H = \{ \text{pos n} \}$

Figure 1: Relative Clause as Adjunct

is unifiable with the right daughter category, characterizes the left daughter (*ie.* relative clause) as an adjunct.

However, we have to explain where this **dep** feature comes from. In Japanse, a relative clause has a tense form at its end. Thus we assume that tense form is categorized as a verb which takes a verb category as its complement. Furthermore, we assume that for every tense form (*e.g.* *-ta* and *-ru*), there are two different lexical entries with the same phonetic form, which we call the *prenominal* tense form and the *assertive* tense form [2]. The former type is assumed to be the ending form of relative clauses, and has a non-null value in **dep** feature. This value is bound with the head noun of the relative clause [3].

## 4 Types of Relative Clauses

Relative clauses can be classified into the following three types in terms of their structural characteristics (where $\phi$ marks a gap):

1. **SS: Simple Sentence**
   The relative clause consists of a simple sentence, and the gap plays a role in that sentence, *e.g.*

   (3) [ Naomi -ga   $\phi$   yom -da ]  hon
        NOM ACC read PAST book
   'the book which Naomi read'

2. **ES: Embedded Sentence**
   The relative clause includes an embedded sentence in which the gap plays a role, *e.g.*

   (4) [ Naomi -ga   [ Taro -ga   $\phi$   yom -da ]
        NOM        NOM ACC read PAST
   -to   sinzitei -ru ]  hon
   CMPL believe PRES book
   'the book which Naomi believes Taro read'

3. **RC: Relative Clause**
   The relative clause includes a relative clause in which the gap plays a role, *e.g.*

   (5) [ [ $\phi$   $\phi$   kak -ta ]  hon -ga
          NOM ACC write PAST book NOM
   yoku ure -ta ]  sakka
   well sell PAST writer
   'the writer whose book sold well'

---

[2]There are several reasons why we can assume there are two different lexical entries for tense with the same phonetic form [9].

[3]The *assertive* tense form is assumed to be the ending form of assertive clauses, and has a null value in **dep** feature.

Table 1: Acceptance of Relative Clause

| Grammatical Role | SS | ES | RC |
|---|---|---|---|
| Subject | o | o | o |
| Object | o | o | ? |
| Adjunct 1 | o | o | ? |
| Adjunct 2 | * | * | * |

Table 2: Simplification of Judgement

| Grammatical Role | SS | ES | RC |
|---|---|---|---|
| Subject | o | o | o |
| Object | o | o | o |
| Adjunct 1 | o | o | o |
| Adjunct 2 | * | * | * |

As Inoue[5, 6] pointed out, the acceptability of the relative clauses varies depending on their types shown above and what roles are played by the gaps.

The grammatical roles other than subject and object can be classified into the following two types depending on the acceptability of relative clauses shown in Table1, where, o, ? and * represent our subjects' judgements are acceptable, marginal, and unacceptable, respectively. :

**Adjunct 1** : This includes 'locative'(marked by *ni*, *wo* and *de*), 'goal'(marked by *ni* and *he*) and 'source'(marked by *kara*).

**Adjunct 2** : This includes 'instrumental'(marked by *de*), 'reason'(marked by *kara*) and 'comparative'(marked by *yori*).

In order to simplify the formalization, we will regard the marginal cases as acceptable, and the result is shown in Table 2. This simplification, however, may be too permissive, because usually in the RC clauses are acceptable only when the gap's role is subject.

## 5 Formalization

In this section we will formalize the syntactic structure of the relative clause in Japanese. First we will formalize the SS type relative clause. Then we formalize the distribution of the **slash** feature. And finally, we examine the other cases, that is, those in which the gap is derived in either an embedded sentence or a relative clause.

### 5.1 Simple Relative Clause

As discussed in section 3, a relative clause is regarded as an adjunct. Thus, the gap should be bound with the **dep** feature value of the *prenominal* tense category.

As shown in the last section, gaps are not restricted to subjects or objects. Optional Adjunct 1 categories ('locative','goal' and 'source') can be a gap. Thus, for simplicity, we here assume that Adjunct 1 type roles are also designated in the value of **subcat** feature.
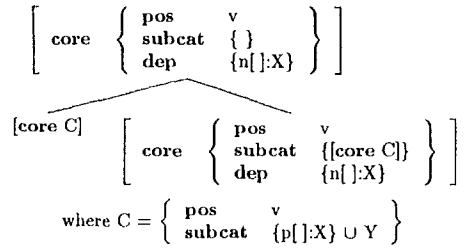


Figure 2: Binding of Dep with Subcat

Furthermore, we assume that a gap is *directly* bound with **dep** feature, instead of being bound with an element of **slash** which is in turn bound with **dep**. We call this hypothesis the *Direct Binding Hypothesis* (that is, **subcat** element can be *directly* bound with **dep**). We will discuss the correctness of this assumption in section 6.

This relationship is expressed in the lexical entry for *prenominal* tense category, eg. *-ta*, *-ru*, and *-i*. The structure is shown in Figure 2, where the head daughter is a *prenominal* tense category. For example, *kak* 'write' is assumed to take both subject and object complements and have the following lexical entry as follows:

(6) v[core {subcat {p[gr subj]:X, p[gr obj]:Y}}]

Then, a relative clause *kak-ta* 'wrote' corresponds to the following categories:

(7) a. v[core {subcat { }, dep { n[gr subj] }}]
    b. v[core {subcat { }, dep { n[gr obj] }}]

Thus for *kak-ta* **X** (where **X** is a noun), there are at least two possibilities in the interpretation of the role of **X**, that is, either subject or object.

### 5.2 Distribution of Slash

As sentence (5) shows, there is a so-called long-distance dependency phenomenon in Japanese. This phenomenon is described by using the **slash** feature as in HPSG. We will characterize the distribution of **slash** in the following three sections.

#### 5.2.1 SLASH INTRODUCTION

Lexical entries with non-null value in **slash** feature are introduced by the following lexical rule:[4]

**Subcat–Slash Lexical Rule:**
    [subcat { p[ ]:X } ∪ S ]
    ⇒ [subcat S, slash { n[ ]:X }]

The number of **slash** elements introduced by this rule can be at most one, a fact which we explain more precisely in the next section.

---
[4]There are some restrictions on the application of this rule. For example, an agglutinated complement cannot become a slash value.

For example, *kak* 'write' will have the following lexical entries other than (6) by this lexical rule:

(8) a. v[core {subcat {p[gr subj]:X}},
    slash {n[gr obj]:Y}]
   b. v[core {subcat {p[gr obj]:Y}},
    slash {n[gr subj]:X}]

### 5.2.2 SLASH INHERITANCE

Each **slash** inheritance is constrained by means of Binding Feature Principle. Here, again, we assume that the number of **slash** elements of each category can be at most one.

In the RC type relative clause, the gap is inherited from the innermost embedded clause to the outermost relative clause shown in Figure 3.
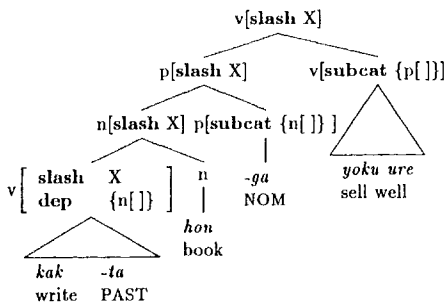


Figure 3: Example of Slash Inheritance

### 5.2.3 SLASH BINDING

A long-distance inherited gap can be bound with the dep of the *prenominal* tense category, and finally bound with the head noun of the relative clause as shown in Figure 4 (where only core features of the mother and the head are shown). For example, if the left daughter category corresponds to *kak-ta hon-ga yoku ure* 'book which $\phi$ wrote sells well'), and the right daughter (*i.e.* the head) to the *prenominal* tense form *-ta*, then the mother, *kak-ta hon-ga yoku ure-ta*, will be described as follows:

$$\text{v[core \{dep \{ n[ ] \}\}]}$$

It should be noted that we do not necessarily need a specific phrase structure rule for the slash binding with the dep [9], though it may be natural that we assume to have such a phrase structure rule as shown in Figure 4.

### 5.3 Complex Relative Clause

In both ES and RC type relative clauses, the Binding Feature Principle specifies that a **slash** element which is introduced by the Subcat-Slash Lexical Rule is inherited and finally bound with the dep value of the *prenominal* tense category in the outermost relative clause. Figure 5 shows an example of such a structure.
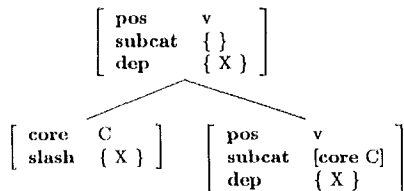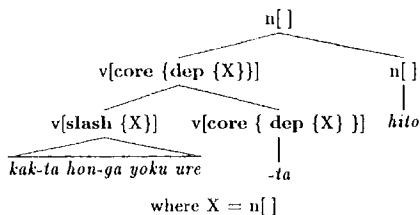


Figure 4: Slash Binding



where X = n[ ]

Figure 5: Complex Relative Clause

## 6 Constraints on the Number of Slash Elements

As the consequence of the constraints on **slash** shown above, an Adjunct 2 type role cannot construct any acceptable relativization, because it cannot introduce an element of **subcat** in any lexical entries.

As discussed in section 3, we assume the relative clause of Japanese language is an adjunct. Thus in the relative clause, the value of the dep feature is bound with the gap. And according to the *Direct Binding Hypothesis* which we proposed in section 5, a **subcat** element can be *directly* bound with the dep. Here, we will discuss why this hypothesis is correct. Furthermore, we will show that if we adopt the *Direct Binding Hypothesis*, the number of slash elements can be at most one.

If we don't accept this hypothesis, the only alternative is to assume that the gap is bound only with an element of slash. That is, if an element of subcat cannot be bound directly with the value of the dep, the gap should be initially in the value of **slash**, and should then be bound with the element of dep. We will call this hypothesis the *Indirect Binding Hypothesis* (that is, subcat element can be *indirectly* bound with the dep through a lexical rule).

First consider the following sentence:

(9) [[ kak  -ta ] hon  -wo syuppan-si -ta ]
    write PAST book ACC publish    PAST
    syuppansya -ga  yuumei -ni   nar
    publisher   NOM famous GOAL become
    -ta ] gakusya
    PAST scholar
    'the scholar such that the publisher that published his book became famous'

This is acceptable to everyone. If we accept the *Indirect Binding Hypothesis*, the number of slash elements would be at least two, because in the above, every relative clause other than outermost one has at least two slash elements.

However, consider the following structures (where $i$ and $j$ are either 1 or 2, and $\phi_1$ and $\phi_2$ are assumed to be bound with $N_1$ and $N_2$, respectively):

(10) a. Double Relative Clause
$[_s [_s \dots [_s \dots \phi_i \dots ]N \dots [_s \dots \phi_j \dots ]N \dots ]N_2 \dots ]N_1$

b. Double Gap in one Relative Clause
$[_s [_s [_s \dots \phi_i \dots \phi_j \dots ]N \dots ]N_2 \dots ]N_1$

c. Embedded Sentence in Relative Clause
$[_s [_s \dots \phi_1 \dots [_s \dots \phi_2 \dots ]\text{-to} \dots ]N_2 \dots ]N_1$

The following are example sentences with the above structures:

(11) a. Double Relative Clause (10.a)
$[[[\phi_1$ kak    -ta ]   e        -ga    $[\phi_2$ mae-ni
paint PAST picture NOM      before
tor  -ta ]  syasin -ni  nitei     -ta ]
take PAST photo DAT resemble PAST
syasin-ka$_2$   -ga   sin -da ]  gaka$_1$
photographer NOM die PAST painter
'the painter such that the photographer died
whose photo taken before resembled the
picture that he painted'

b. Double Gap in one Relative Clause (10.b)
$[[[\phi_1\phi_2\phi_3$ kaka  -se    -ta ]  gakusya$_2$ -ga
write make PAST scholar   NOM
sin -da ]  hon$_3$ -ga   yuumei -ni
die PAST book NOM famous DAT
nar      -ta ]  hensyusya$_1$
become PAST editor
'the editor who made the scholar who died
write a book that became famous'

c. Embedded Sentence in Relative Clause (10.c)
$[[\phi_1[\phi_2$ syuwai-si -ta ]  -to     kak  -ta ]
take-bribe PAST CMPL write PAST
seizika$_2$  -ga  zisatu-si   -ta ]  kisya$_1$
politician NOM kill-himself PAST journalist
'the journalist who wrote a report that the
politician, who killed himself, took a bribe'

Though the sentence (11.c) is better than the others, these are all unacceptable. Taking *Indirect Binding Hypothesis*, the number of slash elements for the above is 3, 3, and 2, respectively. It follows that in terms of the number of slash elements, we cannot explain the difference of the acceptability of the above structures.

However, we accept the *Direct Binding Hypothesis*, sentence (9) needs only one slash element, and sentences (11) all need at least two. Thus the *Direct Binding Hypothesis* is better than the *Indirect Binding Hypothesis*. And adopting this hypothesis leads

to the concusion that the number of slash elements should be at most one.

## 7   Concluding Remarks

We have described a grammatical formalization of the relative clause in Japanese. Based on the JPSG framework, this formalization is characterized in terms of constraints, that is, relationships among the grammatical features in a phrase structure and information in the lexical entries. In this paper we proposed the *Direct Binding Hypothesis*, and showed that the number of slash elements should be at most one in Japanese.

## References

[1] Gunji, T. : *Japanese Phrase Structure Grammar: A Unification-based Approach*, D. Reidel, 1987.

[2] Harada, Y., Gunji, T., Miyoshi, H., Sirai, H. and Hasida, K. : JPSG--A Phrase Structure Grammar for Japanese, *Advances in Software Science and Technology*, Vol. 1, No. 1, 1989, pp. 1-15.

[3] Hasida, K. : Conditioned Unification for Natural Language Processing, *Proc. 11th Int. Conf. on Comput. Linguistics*, 1986, pp. 85–87.

[4] Hasida, K. and Sirai, H. : Conditioned Unification, *Comput. Softw.*, Vol. 3, No. 4 (1986), pp. 28–38 (in Japanese).

[5] Inoue, K. : *Transformational Grammar and Japanese language*, Taishukan, 1976 (in Japanese).

[6] Inoue, K. : *Grammar Rules of Japanese language*, Taishukan, 1978 (in Japanese).

[7] Pollard, C.J. and Sag, I.A. : *Information-Based Syntax and Semantics, Vol. 1: Fundamentals*, CSLI Lecture Notes No. 12, Center for the Study of Language and Information, Stanford University, 1988.

[8] Pollard, C.J. and Sag, I.A. : *Information-Based Syntax and Semantics, Vol. 2*, to be published.

[9] Sirai, H. : *Relative Clause and Adnominal Clause in Japanese*, in preparation.

[10] Tsuda, H., Hasida, K. and Sirai, H. : JPSG Parser on Constraint Logic Programming, *Proc. of 4th ACL European Chapter*, 1989, pp. 95–102.

# Abstract in Japanese

本論文は、JPSG理論に基づいて、日本語の関係節化における統語制約について述べたものである。

JPSG理論では、HPSG理論と同様に、いろいろな情報を担う素性から範疇が構成されていると考える。例えば品詞を表すpos、どういう補語を取るかを表すsubcat、どのような範疇に対する付加語となるかを表すdep、空所の存在を表すslashなどの素性によって、それぞれの範疇が担う情報を記述する。

また、日本語には、補語構造、付加語構造、等位構造の3種類の句構造図式があると考える。そして、それぞれの図式に共通に適用される素性原理と、どのような情報をそれぞれの語彙項目が担っているかということから、さまざまな言語現象を説明しようとする。

本論文では初めに、関係節を名詞句に対する付加語として位置づける。これは、関係節に対応する範疇が、dep素性の値として名詞句と単一化しうる要素をもっているとする事で特徴づけられる。そして、このdep素性は、関係節を構成する動詞句の時制辞である「た」や「る」の語彙項目に記述されているとする（ただし、言い切りの文に現れる時制辞とは別の語彙項目と考える）。すると、どのような要素がこのdep要素と単一化することができるかが問題となる。

よく知られているように、日本語では、主語や目的語だけではなく、位置格や目標格、源泉格なども関係節化される。従って我々は、このような要素も、動詞のsubcat素性の値として記述されると考える。しかし、道具格や理由格、比較格は関係節化されないので、これらはsubcat要素には含めないとする。

ここで、関係節を次のような3種類に分けて解析する。

**SS型（単文）** 関係節が単文である場合
　　例: 奈緒美が読んだ本
**ES型（埋め込み文）** 補文標識「と」を伴う埋め込み文中に空所がある場合
　　例: 奈緒美が太郎が読んだと信じている本
**RC型（関係節）** 関係節中に関係節があり、空所がその埋め込まれた関係節にある場合
　　例: 書いた本がよく売れた作家

関係節がSS型の場合、時制辞を除いた関係節の動詞句（「奈緒美が読んだ本」であれば「奈緒美が読ん」）のsubcatの要素の一つが時制辞（例では「だ」）のdep要素と単一化すると考える。これを直接束縛仮説と呼ぶ。

それに対し、ES型やRC型の関係節では、いわゆる長距離依存現象があるため、HPSG同様、slashを用いて記述する。

**slashの導入:** Subcat-Slash語彙規則により、subcatの一つの要素がslashの要素となり得る。

**slashの伝播:** 束縛素性原理に従う。即ち、どの局所（二分）木においても、親範疇のslash素性の値は、左右の娘範疇のslash素性の値を合わせたものから、もしもslash素性が何か他の素性の値と束縛する場合はその要素を除いたものに等しいとする。

**slashの束縛:** 関係節では、slash要素の値が時制辞のdep要素と束縛しうる。長距離依存の場合は、埋め込み文や関係節中の空所が外側の関係節の時制辞のdep要素と束縛し、それが最終的に主名詞と束縛すると考える。

ここで一つ問題となるのは、SS型の関係節に対してもES型やRC型の関係節と同様にslashを用いた説明が可能である、ということである。これを間接束縛仮説と呼ぶ。これによれば、本来は動詞のsubcat要素であったものがSubcat-Slash語彙規則によりslash要素となり、それが時制辞のdep要素と束縛する事になる。つまり、subcat要素がdep要素と間接的に束縛する。

しかし、ここで次のような言語現象がみられる。それは、

(1) [[[φ₁ 書いた] 本を出版した] 出版社が有名になった] 学者₁

は許容可能な文であるのに対し、

(2) a. [[[φ₁ 描いた] 絵が [φ₂ 前に撮った] 写真に似ていた] 写真家₂ が死んだ] 画家₁

b. [[[φ₁φ₂φ₃ 書かせた] 学者₂ が死んだ] 本₃ が有名になった] 編集者₁

c. [[φ₁[φ₂ 収賄した] と書いた] 政治家₂ が自殺した] 記者₁

はいずれも許容できない文であるという言語現象である。(1)も(2)もどちらも、空所が関係節の主名詞に長距離依存する、という例文である。

この事実は、間接束縛仮説によっては説明できない。しかし、直接束縛仮説を採れば、文(1)が1個のslash要素を必要とするのに対し、文(2)はいずれも2個以上のslash要素を必要とすることから説明可能である（2a,bは3個、2.cは2個必要である）。さらに、直接束縛仮説を認めれば、日本語では許されるslash要素の個数は高々1個であることが導かれる。