# A Multi-Modal Combinatory Categorial Grammar Analysis of Japanese Nonconstituent Clefting

YUSUKE KUBOTA[†]
The Ohio State University
kubota@ling.osu.edu

E. ALLYN SMITH[†]
The Ohio State University
esmith@ling.osu.edu

ABSTRACT. Despite the notion that clefting is a cross-linguistic constituency test, Japanese allows some nonconstituent exceptions. There is, however, a certain restriction on the degree of flexibility; some constituents are more tightly connected (and thus less likely to be separated by clefting) than others. We refine Kubota and Smith's (2006) CCG account in terms of Multi-Modal CCG (Baldridge, 2002): finer-grained modal control provides a means for capturing different degrees of connectedness between an argument and its functor. We then demonstrate how a MMCCG system that finds independent motivation from syntactic complex predicate data interacts with a simple analysis of clefting to account for the full range of clefting patterns. This in turn suggests that what seems to pose problems for a simple analysis of a given phenomenon (clefting) can be overcome once interactions with other phenomena are taken into account.

## 1 Introduction

Multi-Modal Combinatory Categorial Grammar (MMCCG) (Baldridge, 2002) brought together two strands of research in categorial grammar: Combinatory Categorial Grammar (CCG), a more linguistically-oriented variant that has entertained a wider range of empirical applications and Type-Logical Grammar (TLG), a more formally-oriented variant whose logical properties are better understood. Accounting for natural language data, both cross-linguistically and across phenomena within a single language, is crucial for developing a formal theory of natural language. While crosslinguistic work already exists for MMCCG (Baldridge, 2002), the in-depth description of

[†]The authorship of this paper is fully joint; the authors are listed alphabetically.



Proceedings of the Twelfth ESSLLI Student Session
Ville Nurmi and Dmitry Sustretov (editors)

a single language has yet to be undertaken, to the best of our knowledge (and is unfortunately very rare in categorial grammar as compared to other grammatical theories). Thus, this paper takes a first step in that kind of investigation by giving a detailed analysis of the cleft construction in Japanese within a larger theory of grammar of Japanese that handles scrambling and complex predicates. We believe that this kind of work represents a truly interdisciplinary study of logic and language, wherein implications of empirical data are seriously taken into account in theory development through the process of modelling complicated interactions of linguistic phenomena explicitly within a formal theory.

## 2 Data

In Japanese, a cleft sentence is formed by topicalizing a sentence (marked by the topicalizer *wa*, which in turn requires the nominalizer *no*) and combining it with the focused element (an argument or an adjunct missing from the topicalized sentence) followed by the copula *da*.[1]

(1) [ Ken ga $t_i$ kat-ta] no wa *sono hon* *(o)*$_i$ da.
Ken NOM buy-PAST NMLZ TOP that book ACC COP
'It is that book that Ken bought.'

In (1) the object *sono hon o* is missing, appearing instead in the position immediately preceding the copula. Just as in other languages, these cleft sentences are truth-conditionally equivalent to simple sentences but differ in their information structure depending on what is clefted.

In addition to simple constituent clefts such as those in (1), Japanese allows nonconstituent clefts as in (2), as was first noted by Koizumi (1995).[2]

(2) a. [ Ken ga $t_i$ $t_j$ barasi-te simat-ta] no wa *Mari ni*$_i$
Ken NOM disclose EMPH-PAST NMLZ TOP Mari DAT
*sono himitu o*$_j$ da.
that secret ACC COP
lit. 'It is to Mari the secret that Ken (inadvertently) disclosed.'

b. [$t_i$ $t_j$ $t_k$ Barasi-te simat-ta] no wa *Ken ga*$_i$ *Mari ni*$_j$
disclose EMPH-PAST NMLZ TOP Ken NOM Mari DAT
*sono himitu o*$_k$ da.
that secret ACC COP
lit. 'It is Ken the secret to Mari that (inadvertently) disclosed.'

[1] The use of italics in examples indicates the focal position, while the use of brackets indicates the topic position (the focus/topic division roughly corresponds to new/old information distinction); brackets and traces appear for expository ease.

[2] It should be noted that there are speakers who do not accept sentences of this sort (Kizu, 2005). The judgements of the sentences reported here are those of the native-speaking author of this paper.

In these examples, multiple arguments have been clefted together. It is also possible to have multiple adjuncts or argument/adjunct pairs in the focal position. For relevant data, see Kubota and Smith (2006).

Furthermore, the order of elements in the focal position is flexible (that is, if the orders of the accusative and dative objects are switched in (2a), that will still yield a grammatical sentence). This is presumably related to the fact that word order is relatively free in Japanese. That is, Japanese is a verb-final language but allows for scrambling of arguments of the verb. Thus, both sentences with SOV and OSV orders are grammatical.

But there are also some limitations on clefting. (3a) is a case involving a complex predicate construction with the *-te* morphological marking on the embedded verb.[3] Essentially, the ungrammaticality of this example is due to the fact that the embedded verb (*yon-de* 'read') and the matrix verb (*morat-ta* (benefactive)) are separated from one another. Similarly, adjectives modifying nouns cannot be clefted because they cannot be split from those nouns, as in (3b).

(3) a. * [ Morat-ta] no wa *Ken ga* *Mari ni* *sono hon*
BENEF-PAST NMLZ TOP Ken NOM Mari DAT that book
*o* *yon-de* da.
ACC read-MKR COP
lit. 'The thing that was done for the benefit of somebody was that Ken had Mari read that book for him.'

b. * [ $t_i$ Hon o Taroo ga yon-da] no wa *nagai*$_i$ da.
book ACC Taro NOM read-PAST NMLZ TOP long COP
intended: lit. 'It is long that Taro read a book.'

## 3 Kubota and Smith's (2006) analysis of nonconstituent clefting

We now review the previous analysis of Japanese nonconstituent clefting by Kubota and Smith (2006) (K&S) in CCG. Essentially, in K&S's analysis, argument clusters that appear in the focal position of sentences like those in (2) are treated as constituents, employing the technique familiar from the treatment of nonconstituent coordination (Dowty, 1988; Steedman, 1996).

K&S make use of basic CCG combinatory rules of Function Application (FA), Type-Raising (TR) and Function Composition (FC).[4] For lexical entries, K&S assume that each verb has a separate entry for each possible order in which it takes its arguments (thus, a ditransitive verb is assigned

[3] We call this construction the '*-te* form complex predicate'. In (3a), the morpheme appears in the allomorph *-de* with voicing on the initial consonant.

[4] K&S use the Lambek style slash notation. We depart from this and adopt the 'result leftmost' notation that is more commonly adopted in the literature of CCG.

eight separate lexical entries). In addition to this assumption, the following lexical entries for function words are posited:

(4) a. no: $(S\left[\begin{smallmatrix}+N\\-T\end{smallmatrix}\right]\backslash\$)\backslash(S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash\$)$ c. da: $(S_{[-T]}\backslash X)\backslash(S_{[+T]}/X)$
b. wa: $(S\left[\begin{smallmatrix}+N\\+T\end{smallmatrix}\right]\backslash\$)\backslash(S\left[\begin{smallmatrix}+N\\-T\end{smallmatrix}\right]\backslash\$)$

The nominalizer *no* and the topic marker *wa* are identify functions over S-rooted categories (using the \$ convention of CCG (Steedman, 2000)). The features *N* (nominalized) and *T* (topicalized) are binary features governing the distribution of these function words (i.e., the feature specifications of these words ensure that a sentence without *no* and *wa* is an ungrammatical Japanese cleft sentence). The sentence-final copula *da* plays a pivotal role in putting together the topicalized and focused elements by changing the directionality in which the focused element looks for its argument.

K&S's derivation for the argument-cluster cleft sentence in (5) appears in (6), where the different parts of the derivation are split for readability, with the third piece showing how the first two ultimately combine:

(5) [Ken ga watasi-ta] no wa *sono hon o Mari ni* da.
Ken NOM give-PAST NMLZ TOP that book ACC Mari DAT COP
lit. 'It is that book to Mari that Ken gave.'

(6)
Ken ga: $NP_n$; watasi-ta: $S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash NP_a\backslash NP_d\backslash NP_n$; no: $(S\left[\begin{smallmatrix}+N\\-T\end{smallmatrix}\right]\backslash\$)\backslash(S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash\$)$; wa: $(S\left[\begin{smallmatrix}+N\\+T\end{smallmatrix}\right]\backslash\$)\backslash(S\left[\begin{smallmatrix}+N\\-T\end{smallmatrix}\right]\backslash\$)$
Ken ga watasi-ta: $S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash NP_a\backslash NP_d$ (<)
no wa: $(S\left[\begin{smallmatrix}+N\\+T\end{smallmatrix}\right]\backslash\$)\backslash(S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash\$)$ (FC)
$S\left[\begin{smallmatrix}+N\\+T\end{smallmatrix}\right]\backslash NP_a\backslash NP_d$ (<)

sono hon o: $NP_a$ ⇒ $S_{[+T]}/(S_{[+T]}\backslash NP_a)$ (TR); Mari ni: $NP_d$ ⇒ $(S_{[+T]}\backslash NP_a)/((S_{[+T]}\backslash NP_a)\backslash NP_d)$ (TR)
sono hon o Mari ni: $S_{[+T]}/((S_{[+T]}\backslash NP_a)\backslash NP_d)$ (FC); da: $(S_{[-T]}\backslash X)\backslash(S_{[+T]}/X)$
$S_{[-T]}\backslash((S_{[+T]}\backslash NP_a)\backslash NP_d)$ (<)

Ken ga watasi-ta no wa: $S\left[\begin{smallmatrix}+N\\+T\end{smallmatrix}\right]\backslash NP_a\backslash NP_d$; sono hon o Mari ni da: $S_{[-T]}\backslash((S_{[+T]}\backslash NP_a)\backslash NP_d)$
$S_{[-T]}$ (<)

The first part of the derivation shows how the topicalized sentence is formed: FA combines the verb (the entry yielding the OSV order in simple sentences) with the subject, where the object remains unsaturated.[5] In the focal position, both of the missing arguments are first type-raised and the resultant categories are combined via FC, yielding a functor that is looking for the rest of the sentence (namely, the topicalized portion) to become a complete sentence. However, it is looking for this argument in the wrong direction, namely, to the right. The copula crucially comes into play here and flips the directionality of the slash, causing it to look to the left, as seen at the last step of the second part.

[5]The derivation in (6) makes use of FC in combining *no* and *wa*, but there is a fully equivalent derivation involving only FA as well.

At this point, we have seen how K&S's system handles nonconstituent clefting via the interaction of TR and FC. But there are a few points that remain unsatisfactory in this account. Given the flexibility of CCG, that analysis predicts that any string of words that can occur on the leftmost edge of a sentence can be clefted. Thus, it overgenerates sentences such as those in (3) from section 2. The following shows that K&S's analysis incorrectly licenses the ungrammatical sentence in (3b):

(7)

$$\dfrac{\dfrac{\dfrac{\dfrac{\text{hon o}}{NP_a}}{NP_a\backslash(NP_a/NP_a)}\text{TR}\qquad\dfrac{\dfrac{\text{Taroo ga}}{NP_n}\quad\dfrac{\text{yon-da}}{S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash NP_a\backslash NP_n}}{S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash NP_a}<}{S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash(NP_a/NP_a)}\text{FC}\qquad\dfrac{\dfrac{\text{no}}{(S\left[\begin{smallmatrix}+N\\-T\end{smallmatrix}\right]\backslash\$)\backslash(S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash\$)}\quad\dfrac{\text{wa}}{(S\left[\begin{smallmatrix}+N\\+T\end{smallmatrix}\right]\backslash\$)\backslash(S\left[\begin{smallmatrix}+N\\-T\end{smallmatrix}\right]\backslash\$)}}{(S\left[\begin{smallmatrix}+N\\+T\end{smallmatrix}\right]\backslash\$)\backslash(S\left[\begin{smallmatrix}-N\\-T\end{smallmatrix}\right]\backslash\$)}\text{FC}}{S\left[\begin{smallmatrix}+N\\+T\end{smallmatrix}\right]\backslash(NP_a/NP_a)}<$$

$$\dfrac{\dfrac{\dfrac{\text{nagai}}{NP/NP}}{S_{[+T]}/(S_{[+T]}\backslash(NP/NP))}\text{TR}\qquad\dfrac{\text{da}}{(S_{[-T]}\backslash X)\backslash(S_{[+T]}/X)}}{S_{[-T]}\backslash(S_{[+T]}\backslash(NP/NP))}<\qquad\qquad\dfrac{\dfrac{\text{hon o Taroo ga yon-da no wa}}{S\left[\begin{smallmatrix}+N\\+T\end{smallmatrix}\right]\backslash(NP_a/NP_a)}\quad\dfrac{\text{nagai da}}{S_{[-T]}\backslash(S_{[+T]}\backslash(NP/NP))}}{S_{[-T]}}<$$

A related issue is the way scrambling is handled. As mentioned above, K&S posit multiple lexical entries for each verb, taking arguments in different orders (this is necessary not only for licensing all scrambled orders of simplex sentences but also for licensing all the grammatical patterns of nonconstituent clefting). While this may be the simplest solution for flexible word order, it is not particularly elegant and it leads to a peculiar consequence in terms of overgeneration: the derivation above, incorrectly licensed in OSV order, is ruled out in SOV order. As we will see in the next section, these inadequacies can be overcome by adopting a slight modification of the system within Multi-Modal Combinatory Categorial Grammar.

# 4 A Multi-Modal Analysis

Multi-Modal Combinatory Categorial Grammar (MMCCG) (Baldridge, 2002) retains the basic architecture of CCG while adding slash modalities, thereby gaining finer-grained control over structure-manipulating operations in natural language. Baldridge uses extraction asymmetries and scrambling phenomena in English, Dutch, Turkish, Tagalog and Toba Batak to demonstrate how this system can keep the set of combinatory rules (the grammar) constant while capturing language-specific structure-sensitive properties with lexical specifications. Because MMCCG lexically differentiates the combinatoric possibilities of words in terms of different kinds of slashes, and further imposes restrictions on structural operations by making the combinatoric schemata sensitive to these distinctions, it is exactly what is needed to account for the restrictions found in Japanese nonconstituent clefting.

As we saw in (3a), the embedded and embedding verbs in the *-te* form complex predicate cannot be separated from one another by clefting. The following examples clarify the pattern we find in the complex predicate data:

(8) a. Ken ga Mari ni sono hon o yon-de morat-ta.
Ken NOM Mari DAT that book ACC read-MKR BENEF-PAST
'Ken had Mari read that book for him.'

b. [ Ken ga yon-de morat-ta] no wa *Mari ni*
Ken NOM read-MKR BENEF-PAST NMLZ TOP Mari DAT
*sono hon o* da.
that book ACC COP
lit. 'What Ken had read for him was Mari that book.'

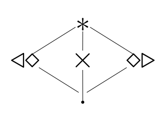
c. * [ Ken ga Mari ni morat-ta] no wa *sono hon*
Ken NOM Mari DAT BENEF-PAST NMLZ TOP that book
*o yon-de* da.
ACC read-MKR COP
intended: 'What Ken had Mari do for him was read that book.'

Here, (8a) is an example of a non-clefted sentence. (8b) is a case in which the embedded accusative object and the matrix dative object are clustered together in the focal position; this example shows that the pattern of clefting in this complex predicate construction is fairly flexible. Basically, as long as the embedded verb (V1) and embedding verb (V2) are not separated from one another, the sentence is grammatical. Example (8c) reveals (perhaps somewhat surprisingly) that it is ungrammatical to cleft the entire embedded verb phrase, but just as in (3a), the ungrammaticality is due to the separation of V1 and V2. In order to account for these data, we crucially distinguish two modes: the normal (or 'scrambling') mode, which is permutative and associative, and the complex predicate mode, which is neither permutative nor right associative, but rather is only left associative.[6] We describe the foundations of such a system below before turning to a demonstration of how this analysis captures the facts above.

### 4.1 The Formal System

The following is the hierarchy of the modes we will employ:[7]

(9)



[6]The existence of these modes is independently motivated to account for further properties of complex predicates, such as scrambling of arguments of the embedded predicate with those of the higher predicate. Unfortunately, space limitations preclude us from discussing these properties here.

[7]Here and elsewhere in this section, we follow the general approach of Baldridge (2002) but differ somewhat in detail in order to assume the minimum theoretical machinery necessary for accounting for the Japanese data. We have no reason not to think that the analysis presented here could be reformulated in Baldridge's system.

The modes are arranged from top to bottom by their permissibility; the top node ($*$) is the least permissive and is neither permutative nor associative in either direction, while the bottom node ($\cdot$) is the most permissive and is both permutative and associative in both directions. The three modes bearing intermediate permissibility each have a single property: ($\triangleleft\diamond$) is left associative, ($\diamond\triangleright$) is right associative, and ($\times$) is permutative.

The distinction between right and left associative modes (which is not present in Baldridge's system) is introduced here in order to distinguish two 'restructuring' operations corresponding to the following binary structural rules in Type Logical Grammar (TLG) (Oehrle, 1998):[8]

(10) a. Right Association $\dfrac{A \bullet_{\diamond\triangleright} (B \bullet_{\diamond\triangleright} C)}{(A \bullet_{\diamond\triangleright} B) \bullet_{\diamond\triangleright} C}$ b. Left Association $\dfrac{(A \bullet_{\triangleleft\diamond} B) \bullet_{\triangleleft\diamond} C}{A \bullet_{\triangleleft\diamond} (B \bullet_{\triangleleft\diamond} C)}$

These TLG rules will be incorporated into the MMCCG system by revising the FC schema in K&S in the following way:

(11) a. $A/_{\diamond\triangleright}B \;\; B/_{\diamond\triangleright}C \;\; \vdash \;\; A/_{\diamond\triangleright}C$ b. $A\backslash_{\triangleleft\diamond}B \;\; C\backslash_{\triangleleft\diamond}A \;\; \vdash \;\; C\backslash_{\triangleleft\diamond}B$

The modality specification here ensures that two functors can be composed only when the modality of each (in addition to the directionality of the slashes) matches.[9] The distinction of left and right associative modes is motivated by empirical evidence: as we will see below, by assigning the left associative mode as the combinatoric mode for the complex predicate formation, the syntactic properties of the *-te* form complex predicate can be neatly captured, including the pattern it exhibits when it interacts with clefting. As for the other combinatoric rules, FA remains unchanged except that it is specified for the least permissive $*$ mode (which ensures that it is applicable to any mode as guaranteed by the convention of rule schema application described in footnote 9). TR requires the following slight revision in order to guarantee that the original combinatory property is preserved after type-raising.[10,11]

[8] We have limited ourselves to modelling the effects of binary structural rules of association and permutation in Oehrle (1998) in CCG, whose linguistic motivations are better understood than unary structural rules. For more recent and detailed discussion of formal characterizations and empirical applications of structural rules in TLG, see Moortgat (1996), Bernardi (2002) and Vermaat (2005).

[9] More precisely, following Baldridge, we assume that combinatoric rules can apply only when the modality specification on the input is more permissive than what is specified in the rule. For example, (11a) is applicable when the modality of the slash of the lefthand element of the input (i.e. what instantiates $A/_{\diamond\triangleright}B$) is the most permissive mode ($\cdot$).

[10] We employ the following convention: any slash without a specified modality is an abbreviation of $/.$ or $\backslash.$, the most permissive mode.

[11] The index $i$ is a variable notation for slash modalities. The purpose of this variable index here is to ensure that the modality specifications match for both slashes when a category is type-raised over another category.

(12) a. $A \ \vdash \ B/_i(B\backslash_i A)$ b. $A \ \vdash \ B\backslash_i(B/_i A)$

In addition to these rules, we introduce one nonlogical unary rule to handle scrambling:[12,13]

(13) a. $A/_\times B/_\times C\$ \ \vdash \ A/_\times C/_\times B\$$ b. $A\backslash_\times B\backslash_\times C\$ \ \vdash \ A\backslash_\times C\backslash_\times B\$$

This enables a functor looking for two categories successively in the same direction (and in the permutative mode) to flip the order of these arguments. This allows each verb to be listed only once in the lexicon with its basic word order, since all other orders can be obtained from that basic entry by successive applications of (13). The following, then, are the lexical entries necessary for the derivations of the relevant examples.[14,15,16]

(14) a. *morat-ta*: $S\backslash NP_n\backslash NP_d\backslash_{\triangleleft\diamond} VP$ c. *nagai*: $NP\backslash_* NP$
b. *yon-de*: $VP\backslash NP_a$

As we have alluded to all along, the benefactive verb *morat-ta* takes its embedded verb argument in the left associative or 'complex predicate' mode. We also see that the prenominal adjective *nagai* 'long' takes its nominal argument in the least permissive mode, indicating the tightest possible relationship between two lexical items.

## 4.2 Accounting for the Data

We begin by demonstrating how the new system handles grammatical cases of nonconstituent clefting with complex predicates, such as (8b).

---

[12]The semantics for these permutation rules can be defined as follows:

(i) a. $A/_\times B/_\times C\$ : \lambda x_0...x_n yz.\boldsymbol{\varphi} \ \vdash \ A/_\times C/_\times B\$ : \lambda x_0...x_n zy.\boldsymbol{\varphi}$
b. $A\backslash_\times B\backslash_\times C\$ : \lambda x_0...x_n yz.\boldsymbol{\varphi} \ \vdash \ A\backslash_\times C\backslash_\times B\$ : \lambda x_0...x_n zy.\boldsymbol{\varphi}$

With these definitions, the straightforward syntax-semantics interface of CCG in K&S's analysis is maintained. For the semantics of other rules and lexical entries and sample derivations illustrating the syntax-semantics interface assumed here, the reader is referred to Kubota and Smith (2006).

[13]An important alternative to this approach is set-based CCG (Hoffman, 1995), as pointed out by an anonymous reviewer. While the analysis of word-order in this paper in terms of a unary permutation rule might be seen as introducing too much flexibility in the grammar, we have opted for this account for the following reasons: (i) it allows one to capture the relevant linguistic generalizations relatively straightforwardly and (ii) it maintains the straightforward syntax-semantics interface of standard CCG in which model-theoretic interpretation is directly obtained from surface composition, which is lost in the multi-set CCG alternative. A detailed comparison of the present proposal and the set-based CCG alternative is beyond the scope of this paper.

[14]$VP$ is an abbreviation of $S\backslash NP_n$.

[15]Entries for *no*, *wa* and *da* are unchanged; slashes will be specified in the $\cdot$ mode.

[16]The verb *morat-ta* is ambiguous between its use as a benefactive predicate and ordinary lexical verb meaning 'receive'. We assume that there is a different ditransitive verb entry for the latter.

(15)

$$
\begin{array}{l}
\text{yon-de} \quad \text{morat-ta} \\
VP\backslash NP_a \quad S\backslash NP_n\backslash NP_d\backslash_{\triangleleft\diamond} VP \\
\overline{S\backslash NP_n\backslash NP_d\backslash NP_a}\ \mathbf{FC} \\
\overline{S\backslash NP_d\backslash NP_n\backslash NP_a}\ \text{Perm} \\
\text{Ken ga} \quad \overline{S\backslash NP_d\backslash NP_a\backslash NP_n}\ \text{Perm} \\
NP_n \\
\overline{S\backslash NP_d\backslash NP_a}\ < \qquad \text{no wa} \quad (S[{}^{+N}_{+T}]\backslash\$)\backslash(S[{}^{-N}_{-T}]\backslash\$) \\
\overline{S[{}^{+N}_{+T}]\backslash NP_d\backslash NP_a}\ <
\end{array}
$$

Here, the V1 and V2 can function compose since the slash modality specifications on the rule subsume those lexically specified for the verbs.[17] From this point on, the cluster of the V1 and V2 effectively functions as a single 'ditransitive verb' looking for arguments of the embedded verb and the higher verb successively. Permutation is then applied to scramble the order of the verbs' arguments in order to combine next with the subject *Ken ga*, and the rest of the derivation proceeds in parallel to the example from K&S in (6).

The next two derivations demonstrate the ability of our system to block the ungrammatical examples seen in (8c) and (3a). In (8c), the dative matrix argument appears linearly adjacent to the matrix verb. Thus, in order for these words to combine, the order of the dative NP and the embedded VP in the lexical specification of the matrix verb needs to be flipped so that the dative NP becomes the first argument that the matrix verb is looking for. However, the permutation rule (13b) cannot be applied here due to the tight connection between the embedded and embedding verbs encoded in the lexical entry of *morat-ta* with the left associative mode ($\triangleleft\diamond$) as illustrated in the following failed derivation:

(16)

$$
\begin{array}{l}
\qquad\qquad \text{morat-ta} \\
\text{Mari ni} \quad S\backslash NP_n\backslash NP_d\backslash_{\triangleleft\diamond} VP \\
NP_d \quad \overline{S\backslash NP_n\backslash_{\triangleleft\diamond} VP\backslash NP_d}\ \mathbf{{*}Perm}
\end{array}
$$

In (3a), on the other hand, it is not permutation that causes the problem. In this case, the fact that the $\triangleleft\diamond$ slash modality for the embedded verb is left associative but not right associative plays a crucial role. In order for (3a) to be derived, the clefted string would have to be analyzed as a constituent that is looking for the matrix verb to become an S. This means that, in the focal position, the embedded VP would have to be type-raised over $S\backslash NP\backslash NP$ so that it could successively function compose with the matrix nominative and dative arguments to result in the desired category. But in order to match the lexical specification on the matrix verb (so that it can eventually combine with it), it would have to type-raise with the $\triangleleft\diamond$ modality. Thus, the embedded VP is type-raised to the category $((S\backslash NP_n)\backslash NP_d)/_{\triangleleft\diamond}(((S\backslash NP_n)\backslash NP_d)\backslash_{\triangleleft\diamond} VP)$, as we see in (17):

[17] The use of FC for forming verb clusters is a standard technique for analyzing complex predicates in CCG (see, for example, the analysis of the Dutch cross-serial dependency construction by Steedman (2000)).

(17)

$$\dfrac{\dfrac{\dfrac{\text{Mari ni}}{NP_d}}{(S\backslash NP_n)/((S\backslash NP_n)\backslash NP_d)}\text{TR} \quad \dfrac{\dfrac{\text{sono hon o yon-de}}{VP}}{((S\backslash NP_n)\backslash NP_d)/_{\triangleleft\diamond}(((S\backslash NP_n)\backslash NP_d)\backslash_{\triangleleft\diamond} VP)}\text{TR}}{}\textbf{*FC}$$

Being in this new category, however, the embedded VP cannot function compose by (11a) with the type-raised matrix dative argument since $\triangleleft\diamond$ isn't right associative, and so the derivation fails, ruling out this sentence which K&S's analyses overgenerates.

Finally, in prenominal modification, the ungrammatical example in (3b) that was overgenerated by K&S's system is also blocked in our analysis:

(18)

$$\dfrac{\dfrac{\dfrac{\text{hon o}}{NP_a}}{NP_a\backslash_{*}(NP_a/_{*}NP_a)}\text{TR} \quad \dfrac{\dfrac{\text{Taroo ga}}{NP_n} \quad \dfrac{\dfrac{\text{yon-da}}{S\backslash NP_n\backslash NP_a}}{S\backslash NP_a\backslash NP_n}\text{Perm}}{S\backslash NP_a}<}{}\textbf{*FC}$$

In this example, in order to derive the topicalized sentence, the noun *hon o* 'book' and the sentence *Taroo ga yon-da* 'Taro read' need to combine via FC. However, due to the mismatch of slash modality, the derivation is blocked (crucially, *hon o* has to first type-raise with the $*$ modality in this example so that it would match the lexical specification of the adjective *nagai* 'long' and ultimately combine with it). Thus, we have seen that the addition of a minimal number of modes enables the structural control necessary to correctly derive all of the grammatical examples while blocking all of the ungrammatical examples.

## 5 Conclusion

Though K&S's CCG analysis of Japanese nonconstituent clefting succinctly captures the essential flexibility of clefting in Japanese, we have seen that further restrictions are necessary to accurately model the full range of data. Adopting MMCCG and positing a few slash modalities independently-motivated to treat complex predicates, the analysis proposed here accurately accounts for all of the data. In addition, these modalities enable the introduction of the permutative rule, which accounts for Japanese scrambling phenomena in a more elegant way than in the previous approach. Given that this analysis effectively describes the interaction between clefting, scrambling, and complex predicates in Japanese, we hope that it will serve as a springboard for further investigations into a more complete and integrated account of Japanese syntax.[18]

[18] As an anonymous reviewer points out, an interesting question for future research is to see whether an account of a wider range of scrambling phenomena is possible in the present setup of MMCCG; it is known that clause-internal scrambling and long-distance scrambling behave differently in terms of phenomena such as quantifier scope. For an overview of scrambling phenomena in Japanese and relevant literature, see Nemoto (1999).

# Bibliography


Baldridge, Jason. 2002. *Lexically Specified Derivational Control in Combinatory Categorial Grammar*. Ph.D. thesis, University of Edinburgh. [Available at `http://homepages.inf.ed.ac.uk/jbaldrid/dissertation/index.html`].

Bernardi, Raffaella. 2002. *Reasoning with Polarity in Categorial Type Logic*. Ph.D. thesis, University of Utrecht. [Available at `http://www.inf.unibz.it/~bernardi/finalthesis.html`].

Dowty, David. 1988. Type raising, functional composition, and non-constituent coordination. In R. T. Oehrle, E. Bach, and D. Wheeler, eds., *Categorial Grammars and Natural Language Structures*, vol. 32 of *Studies in Linguistics and Philosophy*, 153–198. Dordrecht: Kluwer.

Hoffman, Beryl. 1995. *The Computational Analysis of the Syntax and Interpretation of 'Free' Word Order in Turkish*. Ph.D. thesis, University of Pennsylvania.

Kizu, Mika. 2005. *Cleft Constructions in Japanese Syntax*. Houndmills, England and New York: Palgrave Macmillan.

Koizumi, Masatoshi. 1995. *Phrase Structure in Minimalist Syntax*. Ph.D. thesis, MIT.

Kubota, Yusuke and E. Allyn Smith. 2006. Syntax and semantics of Japanese nonconstituent clefting in Combinatory Categorial Grammar. In C. Davis, A. R. Deal, and Y. Zabbal, eds., *Proceedings of the Thrity-Sixth Annual Meeting of the North East Linguistic Society*, 413–426. Amherst, MA: GLSA Publications.

Moortgat, Michael. 1996. Categorial type logics. In J. van Benthem and A. ter Meulen, eds., *Handbook of Logic and Language*, 93–177. Amsterdam: Elsevier.

Nemoto, Naoko. 1999. Scrambling. In N. Tsujimura, ed., *The Handbook of Japanese Linguistics*, 121–153. Oxford: Blackwell.

Oehrle, Richard T. 1998. Multi-modal type-logical grammar. To appear in R. Borsley and K. Borjars eds., *Non-transformational Syntax*. Blackwell.




Proceedings of the Twelfth ESSLLI Student Session
Ville Nurmi and Dmitry Sustretov (editors)

Steedman, Mark. 1996. *Surface Structure and Interpretation*. Cambridge, Massachusetts: The MIT Press.

Steedman, Mark. 2000. *The Syntactic Process*. Cambridge, Massachusetts: The MIT Press.

Vermaat, Willemijn. 2005. *The Logic of Variation. A cross-linguistic account of wh-question formation*. Ph.D. thesis, Utrecht Institute of Linguistics OTS, Utrecht University.