Syntax and Semantics of Japanese Nonconstituent Clefting in Combinatory Categorial Grammar*

Yusuke Kubota and E. Allyn Smith†

The Ohio State University

1. Introduction

In the Japanese cleft construction, strings composed of multiple phrases that apparently do not form constituents can occupy the focus position. This poses a significant challenge to mainstream syntactic frameworks that have the notion of phrase structure as their backbone. In fact, in the Minimalist tradition, there have been three proposals in the recent literature (Koizumi 1999, 2000; Takano 2002; Fukui and Sakai 2003) that introduce different operations to treat this phenomenon. While each of these works is suggestive in several ways, capturing some aspect of this construction within the assumed framework, none of them stand up to the full range of data without running into inconsistent consequences, as we will see below.

Categorial grammar provides a particularly attractive framework for this problem because its independently motivated theoretical assumptions lead to a grammar that naturally licenses the kind of unusual constituents we find in this construction. The goal of the present paper is to develop an analysis of Japanese nonconstituent clefting in Combinatory Categorial Grammar (CCG) (Steedman 1996, 2000b), in which a straightforward and precise syntax-semantics-information structure account of this construction is given. In constructing our analysis, we avail ourselves of only those assumptions and mechanisms that have been proposed in the literature of CCG and that find empirical motivation elsewhere in the grammar of Japanese. We argue that the resultant analysis is simpler and more explicit than any of the previous analyses and that it accounts for all the relevant data while

*We would like to thank the following people for many helpful comments and suggestions: Peter Culicover, Hideki Kishimoto, Bob Levine, Carl Pollard and Craige Roberts. We have also received valuable feedback from the audiences of the syntax discussion group Synners at OSU and NELS 36 at UMass. Needless to say, we are solely responsible for any remaining errors.

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

facing the least number of potential problems or unresolved issues.

2. The cleft construction in Japanese and nonconstituent clefting

2.1. Basic patterns of the cleft construction in Japanese

In Japanese, cleft sentences are formed by combining a topicalized sentence[1] containing a gap with its 'filler' (the missing element) by means of a copular construction as in (1):[2,3]

(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 this sentence, the object *sono hon o* 'that book' of the transitive verb *kat-ta* 'bought' is missing from the topicalized sentence, and that object appears in the position immediately preceding the copula.

Semantically and pragmatically, the element that appears in the position immediately before the copula is the focus (in the sense that it conveys new information) and the gapped sentence marked by the topic marker is, as expected, the topic (in the sense that it is old information) (Martin 1975/1983). Thus, (1) can be an answer to the question *Ken wa nani o kat-ta no?* 'What did Ken buy?', but not to *Dare ga sono hon o kat-ta no?* 'Who bought that book?', nor *Ken wa nani o si-ta no?* 'What did Ken do?'.

The following are examples in which subject (2), indirect object (3), from-PP (4) and adverb (5) are clefted.[4]

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

(3) [Ken ga sono hon o t_i mise-ta] no wa *Mari ni*$_i$ da.
Ken NOM that book ACC show-PAST NMLZ TOP Mari DAT COP
'It is to Mari that Ken showed that book.'

(4) [Ken ga sono hon o t_i toriyose-ta] no wa *Huransu kara*$_i$ da.
Ken NOM that book ACC order-PAST NMLZ TOP France from COP
'It is from France that Ken ordered that book.'

[1] To be topicalizable, a sentence first has to be nominalized. Hence the presence of the nominalizer *no*.

[2] Adjuncts can also be clefted. In that case, there is no gap in the topicalized sentence, at least in the ordinary sense of this term.

[3] Traces and brackets in examples appear solely for expository purposes. Throughout this paper, the focus position is indicated in *italics* and the topic position is indicated in brackets in the examples.

[4] As shown in (2), when the subject of the sentence appears in the focus position, the nominative marker is obligatorily omitted (Takano 2002). The accusative marker is optional as in (1) and other (more oblique) markers are obligatory as in (3) and (4). Similar (but not identical) patterns of deletion of case markers are found in other environments such as the topic position of a simple sentence and the position immediately preceding the genitive marker. We speculate that the distribution of the case markers in the cleft focus position observed here can be accounted for by independently motivated principles governing other case marker deletion patterns.

(5) [Ken ga hon o t_i yomi-oe-ta] no wa *tui sakki*$_i$ da.
Ken NOM book ACC read-finish-PAST NMLZ TOP just a while ago COP
'It is just a while ago that Ken finished reading the book.'

2.2. Nonconstituent clefting

As noted by Koizumi (1999), strings apparently not forming constituents can occupy the focus position in the cleft construction in Japanese. In the following examples, multiple arguments are clefted and appear together in the focus position:

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

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

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

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

Clefting of multiple phrases is not limited to arguments. The following is an example where multiple adjuncts are clefted. In this sentence, the temporal and locative adjuncts *kyonen* 'last year' and *NELS de* 'at NELS' together occupy the focus position:

(10) [Taroo ga t_i t_j happyoo-si-ta] no wa *kyonen*$_i$ *NELS de*$_j$ da.
Taro NOM present-do-PAST NMLZ TOP last-year NELS at COP
'It is last year at NELS that Taro presented (a paper).'

Combinations of arguments and adjuncts are also possible, as in the following example, where the adjunct *kinoo* 'yesterday' and the object *sono hon o* 'that book' are clefted together.

(11) [Ken ga t_i t_j kat-ta] no wa *kinoo*$_i$ *sono hon o*$_j$ da.
Ken NOM buy-PAST NMLZ TOP yesterday that book ACC COP
lit. 'It was yesterday that book that Ken bought.'

2.3. Impossible cases of clefting

As seen above, the Japanese cleft construction exhibits a fairly flexible pattern with respect to the elements that can appear in its focus position. However, it is not the case that any phrase in a sentence can be freely clefted. (12) and (13) are illicit cases of clefting.

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

(13) * [Ken ga Mari ni t_i t_j morat-ta] no wa *sono hon* o_i *yon-de*$_j$ da.
Ken NOM Mari DAT BENEF-PAST NMLZ TOP that book ACC read-MKR COP
intended: 'What Ken had Mari do for him was read that book.'

Example (12) shows that an adjective that modifies a noun cannot be split from the head noun and placed in the focus position alone; (13) shows that an embedded verb that is subcategorized for by the higher verb in a certain type of complex predicate construction cannot be split from the higher verb by clefting.

3. Previous analyses

To date, there are three major proposed analyses of nonconstituent clefting in Japanese (Koizumi 1999, 2000; Takano 2002; Fukui and Sakai 2003).[5] In this section, we briefly review these proposals and point to some problems for each.

3.1. Koizumi (1999, 2000)

Koizumi (2000) analyzes examples of nonconstituent clefting like the following in terms of string vacuous verb raising.

(14) [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.'

Roughly speaking, in his analysis, a verb is first raised up to a higher position c-commanding the VP (or S) constituent and then the null-headed remnant VP (or S) moves to the focus position. Thus, (14) gets analyzed approximately along the following lines:

(15) Ken ga t_i watasi-ta$_j$ no wa [$_{\rm VP}$ sono hon o Mari ni t_j]$_i$ da.

There are several theory-internal and empirical problems for this analysis pointed out in the literature (Takano 2002; Fukui and Sakai 2003; Fukushima 2003), some of which are more convincing than others. Due to space limitations, we will not review these criticisms here. Instead, we note two fairly simple and obvious problems for this approach which, to the best of our knowledge, are neither adequately addressed in Koizumi's (1999, 2000) original work nor discussed in previous critiques.

First, given the possibility of remnant VP movement, and assuming that local scrambling is accounted for under the framework adopted by Koizumi as movement out of the VP, it remains a mystery that completely empty VPs cannot move to the focus position. That is, for a source sentence like *Hon o Taroo ga yon-da* 'Taro read a book', which

[5] Fukushima (2003) also deals with this issue, but unlike him, we do not take it to be the case that nonconstituent clefting is possible only in the presence of numerical classifiers. (His analysis crucially hinges on this assumption.)

involves a completely empty VP node in his analysis, it is not clear why an example such as (16), in which no overt string moves out of the topicalized source sentence, is ruled out.

(16) * [Hon o]$_k$ Taroo ga t_i yon-da$_j$ no wa [$_{\rm VP}$ t_k t_j]$_i$ da.
book ACC Taro NOM read-PAST NMLZ TOP COP
intended: 'It is that Taro read the book.'

Of course, one could make recourse to semantic/pragmatic infelicity to account for the unacceptability of such sentences. If (16) has no focused element, then it does not convey new information. Would such a sentence serve a function in discourse? It is suggested in the literature (Steedman 2000a:663) that there are possible functions of all-topic sentences in certain kinds of contexts. Thus, in order for the pragmatic account to go through, it would have to be independently established that cleft sentences like (16) cannot serve such purposes.

A second point is related to the status of verb raising. As Koizumi admits, ordinary (non-null-headed) VPs cannot move to positions that remnant VPs can move to. Thus, clefting of an ordinary VP is ungrammatical as shown in (17):

(17) * [Taroo ga t_i] no wa [$_{\rm VP}$ kono hon o yon-da]$_i$ da.
Taro NOM NMLZ TOP this book ACC read-PAST COP
intended: 'What Taro did was read this book.'

Following Koizumi (2000), let us assume here that the verb raises to C by overt verb raising. The ungrammaticality of (17) by itself does not directly pose a problem for his analysis, since it can be accounted for by requiring verb raising to be obligatory and further assuming that VPs but not CPs can undergo movement to the focus position. Given this, though, a problem arises as to how to account for the coordination of constituents containing both ditransitive verbs and their direct objects, as in the following:

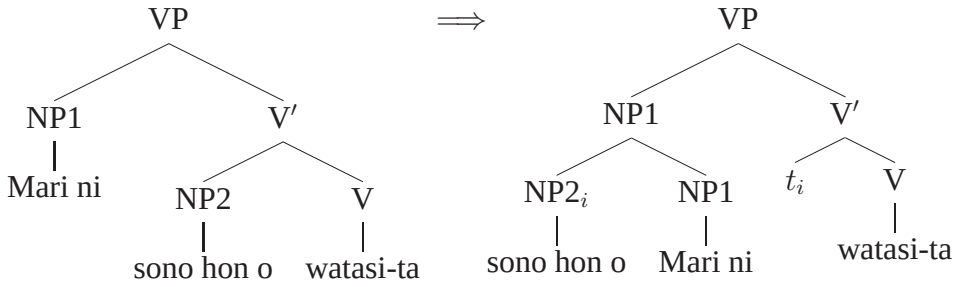
(18) John wa Bill ni [sono hon o kasi]-te, [kono zassi o age]-ta.
John TOP Bill DAT that book ACC lend this magazine ACC give-PAST
'John lent that book and gave this magazine to Bill.'

Assuming obligatory verb raising, there is no node that directly dominates the direct object and the verb that subcategorizes for it in sentences like (18), since the verb raises to a position that c-commands the indirect object. Thus, it is not clear how the coordination of subparts of VPs in (18) is licensed in his analysis, given that the bracketed strings do not form constituents by themselves. Of course, one could make recourse to additional movement operations to create a syntactic node exclusively containing relevant elements, but without independent motivation, such a solution remains non-explanatory.

3.2. Takano (2002)

Takano's analysis of nonconstituent clefting crucially relies on the possibility of adjoining an NP to another NP. That is, under his analysis, sentence (14) is derived by clefting the derived NP constituent consisting of two sub-NPs obtained by the following adjunction operation, which he calls oblique movement:

(19)



In effect, this amounts to saying that the apparent nonconstituent that appears in the focus position is actually a constituent and, moreover, that it is (at least when all of the elements involved are NPs) an NP (in what follows, we call such NPs bipartite NPs).

There are at least two problems with this approach. One involves the coordination of bipartite NPs. Takano is forced to introduce a nonstandard and technically infeasible assumption about coordination to account for cases of coordination of bipartite NPs without verb raising. Space limitations preclude us from giving a detailed discussion on this point. For details, the reader is referred to Kubota and Smith (2005).

The other problem concerns the 'NP-ness' of the purported bipartite NPs. Contrary to the prediction of his analysis, these bipartite NPs do not behave like ordinary NPs. As shown by the following data, ordinary NPs in Japanese can be topicalized as in (20a) but bipartite NPs cannot as in (20b).[6]

(20) a. [Rika ni to Mari ni]$_i$ wa Ken ga t_i kono zassi o watasi-ta.
Rika DAT and Mari DAT TOP Ken NOM this magazine ACC give-PAST
'To Rika and to Mari, Ken gave this magazine.'

b. * [[Sono hon o Rika ni] to [kono zassi o Mari ni]]$_i$ wa Ken ga
that book ACC Rika DAT and this magazine ACC Mari DAT TOP Ken NOM
t_i watasi-ta.
give-PAST
intended: lit. 'That book to Rika and this magazine to Mari, Ken gave.'

The contrast between (20a) and (20b) does not follow from Takano's analysis in which the topicalized phrases in these examples have exactly the same categorical status.

3.3. Fukui and Sakai (2003)

Fukui and Sakai (2003) propose yet another analysis of nonconstituent clefting in Japanese within the framework of Minimalist syntax. According to them, the Japanese cleft is generated not by moving elements from the topicalized sentence to the focus position but by moving a gapped nominalized sentence to the topic position. Thus, (14) is licensed by first

[6] In these examples, coordinated (bipartite) NPs are topicalized in order to avoid the possibility of an irrelevant analysis for apparent bipartite NP topicalization in which a non-topicalized NP is followed by a topicalized NP.

scrambling the two NPs to be clefted to the sentence initial position and then by moving the whole sentence from which these two elements were extracted to the topic position:[7]

(21) [$_S$ Ken ga t_i t_j watasi-ta]$_k$ no wa [sono hon o]$_j$ [Mari ni]$_i$ t_k da.
Ken NOM give-PAST NMLZ TOP that book ACC Mari DAT COP
lit. 'It is this book to Mari that Ken gave.'

There are at least three problems with this approach. Two of the three are due to Fukui and Sakai's (2003) assumption that the movement operation involved in cleft formation is not movement *into* the focus position but movement *out of* it. Thus, these problems are not applicable to Koizumi's or Takano's approach. These involve binding connectivity and a contrast between scrambling and the cleft in terms of the possibility of dislocating conjuncts out of coordinate structures.[8] In the interest of full disclosure, these are potentially problematic for our CCG analysis as well, although it remains to be seen whether these are really problems, given the absence of a complete and elaborate account of Japanese binding and scrambling in CCG.

The other problem for Fukui and Sakai's approach, however, is not applicable to our analysis. Like Koizumi's analysis, this analysis does not account for the fact that sentences like the following, in which no element occupies the focus position, are ungrammatical:

(22) * [Taroo ga hon o yon-da]$_i$ no wa t_i da.
Taro NOM book ACC read-PAST NMLZ TOP COP
intended: lit. 'It is that Taro read the book.'

This sentence would be generated in Fukui and Sakai's analysis by topicalizing the whole sentence without first scrambling anything out of it.[9]

4. A new analysis

Whereas the analyses we have seen in the previous section all exhibit empirical shortcomings, the unified analysis presented in this section avoids these problems by making exclusive use of function application, function composition, and type-raising, along with the lexical specification of the construction-specific items *no*, *wa* and *da* to derive the desired effects.

4.1. The framework

Within Combinatory Categorial Grammar, syntactic categories are specified in the lexicon along with semantic and phonological information. Furthermore, the names of these syntactic categories transparently show what they subcategorize for, so the verb *left*, rather

[7] To be fair, we are making the assumption that the clefted NPs are scrambled. We cannot know whether this is what the authors intended, as they did not comment on how these NPs are displaced from the sentence to be topicalized.

[8] These are discussed in full detail in Kubota and Smith (2005).

[9] Of course, the same caveat as the one noted with respect to our criticism of Koizumi's (2000) analysis applies here. If some pragmatic principle that rules out (22) is independently motivated, this criticism will no longer go through.

than being of type VP would be of type $NP\backslash S$ because it takes an NP as argument to its left to return a S.[10] Given these lexical type specifications, there are several ways in which these items can be combined according to the logical operations available in the grammar. In Japanese, function application (FA), function composition (FC), and type-raising (TR) are all necessary, just as in other languages. The schemas for each of these combinators are given below. For a more detailed introduction to CCG, see Steedman (1996, 2000a).

(23) Function Application

a. A/B: f B: a $\vdash$ A: $f(a)$

b. A: a $A\backslash B$: f $\vdash$ B: $f(a)$

Type-Raising

a. A: a $\vdash$ $B/(A\backslash B)$: $\lambda f.f(a)$

b. A: a $\vdash$ $(B/A)\backslash B$: $\lambda f.f(a)$

Function Composition

a. A/B: f B/C: g $\vdash$ A/C: $\lambda h.f(g(h))$

b. $A\backslash B$: g $B\backslash C$: f $\vdash$ $A\backslash C$: $\lambda h.f(g(h))$

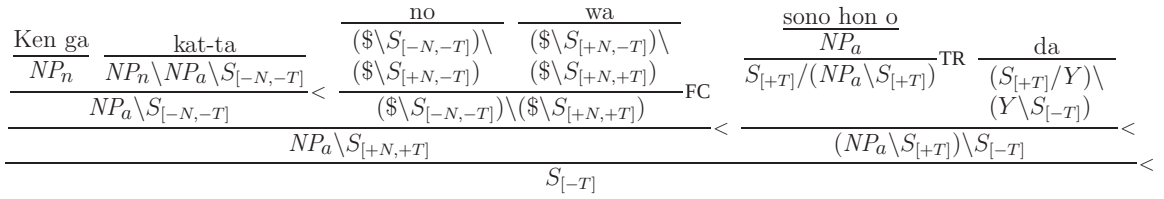
FA is the default means of combining a function with its argument. When functions are instead combined using FC, the resulting string is a 'strange constituent' in that one of its parts would have normally combined with an argument before combining with the other part, but in FC, these two functors combine first and then take the argument later. As in the semantic literature, TR changes the function/argument relationship between two constituents such that a constituent that begins as the argument of some function turns into (via TR) the functor that takes that function as its argument to produce the same result as the one obtained by applying the original function to the original argument.

4.2. Clefting single constituents

The derivation in our analysis for sentence (1) (a case in which the object of the verb is clefted) is given in (24). For the purpose of exposition, the semantic derivations have been separated from the syntactic derivations (and in the case of the first example, presented later) despite the fact that they are in one-to-one correspondence. The fact that the semantic derivations follow the syntactic ones is not indicative of the relative order of components in the grammar; we assume that these are processed in tandem. Also, lambda reductions have not been given separate steps, but rather, all possible conversions have been performed where appropriate in the semantic derivations.

[10] We adopt the Lambek-style slash notation, departing in this respect from the convention of CCG. In the absence of parentheses, the forward slash is left associative and the backward slash is right associative. For example, $NP_a\backslash NP_n\backslash S$ is a shorthand for $NP_a\backslash(NP_n\backslash S)$.

(24) Syntax:



Generally speaking, here we see a topicalized gapped sentence on the left (terminating with *wa*) and the argument that would fill its gap on the right. This argument is type-raised, however, so it becomes the functor that is looking for a gapped sentence as an argument to its right to return a saturated sentence. Once *da* combines with the type-raised constituent, it looks for its argument (the gapped sentence) to its left, where there is indeed a gapped sentence to be found. These then combine to yield a syntactically and semantically well-formed sentence of the language with semantic content that is truth-conditionally identical to its simple sentence counterpart. Note further that the cleft sentence contains exactly the same set of content words (that is, words other than *no*, *wa*, and *da*) as its simple sentence counterpart. The resource sensitivity of the logical system underlying the grammar requires that each expression be used once and only once in the derivation of the sentence.

More specifically, the derivation above relies on the syntactic and semantic information contained in the following lexical specifications for the various markers we find in the Japanese cleft construction:

(25) a. no: $(\$\backslash S_{[-NMLZD,-TOP]})\backslash(\$\backslash S_{[+NMLZD,-TOP]})$: $\lambda f.f$
b. wa: $(\$\backslash S_{[+NMLZD,-TOP]})\backslash(\$\backslash S_{[+NMLZD,+TOP]})$: $\lambda f.\theta'(f)$
c. da: $(S_{[+TOP]}/Y)\backslash(Y\backslash S_{[-TOP]})$: $\lambda f.\rho'(f)$

In (24), the metavariable \$ in the lexical specifications for *no* and *wa* is instantiated as NP_a so that they can take the gapped sentence *Ken ga kat-ta* 'Ken bought' as argument. Similarly, The variable *Y* in the lexical specification for *da* is instantiated as $NP_a\backslash S_{[+T]}$ so that the copula can take the type-raised object as its argument. The markers *no* and *wa* cannot combine with just any category; they require a (possibly gapped) sentential category as specified in their syntactic types. There is, however, no constraint on the number and categories of the gaps, since no constraint is imposed on how the metavariable \$ is to be instantiated. This property will be crucially exploited in the account of multiple constituent cleftng as we will see in the next section.

The binary features NMLZD and TOP are used above to mark whether a constituent has been (syntactically) nominalized or topicalized, and their addition is necessitated by the ungrammaticality of cleft sentences lacking either *no* or *wa* (or both). Generally acknowledged to be a 'topic marker' in Japanese, *wa* is restricted to taking nominalized phrases as arguments. Because its argument in the cleft construction is never a regular nominal (and always a sentential category), *no*'s presence is crucial in licensing *wa*. That is, given that the presence of *no* is necessitated only as a consequence of the syntactic requirement from the topic marker *wa* in our analysis, we can think of these markers as acting as a single unit. This fact is mirrored in the definition for *no*, which changes its argument's

nominal value from [–NMLZD] to [+NMLZD], which we see is the required input to the *wa* function, which in turn takes it from being [–TOP] to [+TOP].[11] In the above derivation, these two words are combined via function composition. While this is not the only derivation that is licensed by the grammar, it nicely highlights the workings of these markers in an intuitively perspicuous way. FC applied to these categories yields the category $(\$\backslash S_{[-N,-T]})\backslash(\$\backslash S_{[+N,+T]})$ corresponding to the string *no wa*. This, in effect, means that by combining with *no wa*, the gapped sentence becomes both nominalized and topicalized. Meanwhile, the clefted material type-raises over a [+TOP] category because of the requirement from the copula, which reflects the fact that this type-raised argument eventually takes a topicalized sentential category as its argument.

Clefting of the subject is also straightforward. Our analysis supposes that verbs are lexically ambiguous between various subcategorizations (i.e. if a subject were clefted, a distinct homophonous verb *kat-ta* would appear in the derivation, of type $NP_a\backslash NP_n\backslash S_{[-N,-T]}$), which is empirically motivated by the argument scrambling found generally in Japanese.[12]

Turning now to the semantics, the 'contentful' lexical items in the clefted and topicalized constituents are given their traditional denotations, and those of the clefting markers (above) are taken to involve information structure. The semantic derivation for sentence (1) is given in (26).

(26) Semantics:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\text{Ken ga}}{k}\quad\dfrac{\text{kat-ta}}{\lambda y\lambda x.bought'(y,x)}}{\lambda x.bought'(k,x)}<\quad\dfrac{\dfrac{\text{no}}{\lambda f.f}\quad\dfrac{\text{wa}}{\lambda f.\theta'(f)}}{\lambda f.\theta'(f)}\text{FC}}{\theta'(\lambda x.bought'(k,x))}<\quad\dfrac{\dfrac{\dfrac{\text{sono hon o}}{\iota x.book'(x)}}{\lambda P.P(\iota x.book'(x))}\text{TR}\quad\dfrac{\text{da}}{\lambda f.\rho'(f)}}{\rho'(\lambda P.P(\iota x.book'(x)))}<}{\rho'(\lambda P.P(\iota x.book'(x)))(\theta'(\lambda x.bought'(k,x)))}<}{bought'(k,\iota x.book'(x))}$$

We follow Steedman (2000a) in the use of his semantic θ' and ρ' markers, wherein they are semantically identity functions, but information-theoretically serve to mark a particular constituent as being part of the theme or rheme.[13] Basically, then, these markers can be thought of as taking truth-conditional content and feeding it to the information structure where it is marked as being old or new in discourse, etc. And in fact, the primary difference between the cleft sentence and its simple sentence counterpart is that the cleft manifests this theme/rheme distinction explicitly in the syntax. Thus, as previously mentioned, they are truth conditionally equivalent, as should be evident from the result of the last step in the derivation in (26). The term that is given at the penultimate step is already the result of

[11] As it is, the treatment of the function words here is construction-specific. We leave open the possibility of integrating the core analysis of the cleft construction proposed here with a more general treatment of these function words in future research.

[12] There may be an analysis of scrambling phenomena in Japanese that is better than positing lexical ambiguity, but because this issue is orthogonal to an analysis of nonconstituent clefting, we have adopted this simplifying assumption at the present time.

[13] These semantic markers should not be confused with the more general syntactic/prosodic θ- and ρ-marking also present in Steedman's article. The necessity of the latter in the grammar of Japanese remains to be shown. In particular, we do not yet know whether Japanese clefting relies on any phonological cues such as pitch accenting or boundary tones to delineate its theme and rheme components.

the semantic derivation, but the step that follows it is included to aid the reader and should not be interpreted as a violation of the one-to-one correspondence between the syntax and semantics.[14]

In this section, we have seen how the derivation of a single-argument cleft works in our analysis and how the lexical definitions of the markers are motivated and contribute to the derivation. When viewed as a whole, the three markers effectively work together to establish a certain 'schema' having slots associated with specific discourse functions. It is important to note, however, that this 'schema' emerges in our analysis solely as a consequence of the interactions of the lexical elements that constitute the cleft construction, rather than being posited as a theoretical primitive as in Construction Grammar. In this sense, our analysis can be seen as reconciling the divide between constructional approaches to grammar and the tradition of formal semantics adhering to strict compositionality by suggesting a way of making sense of the notion of 'schema' in a strictly compositional semantics.

4.3. Nonconstituent clefting analysis

Unlike the example in the previous section, the derivations in this section show clefted elements that are not typically taken to be constituents. Using the same combinators and lexical specifications as above, however, we see that they easily fall out from the general proposal. In our analysis, these nonconstituents are formed via the interaction of TR and FC, where FC is responsible for combining them, as in the analysis of nonconstituent coordination (NCC) in CCG (Steedman 1996) and categorial grammars in general (Dowty 1988).[15] As we have already seen in the case of single constituent clefting, even after TR flips the functor-argument relation between an NP and a gapped sentence looking for it, the former still works as a filler for one gap position in the latter. The trick for the nonconstituent case will be that by converting both of the (clefted) argument NPs to functors by TR, it becomes possible to further combine them into a complex functor that takes the doubly gapped sentence as argument and effectively acts as the two-filler pair that completes it to return a fully saturated sentence. This is what we see in the case of the derivation for (14) in (27).[16]

[14]Procedural metaphors such as 'last' or 'result' are not reflections of the architecture of the grammar.

[15]NCC is also present in Japanese, and though an exact analysis remains to be given, it would require FC and TR to license even the simplest of examples, just as in English. This provides independent motivation for FC and TR in the grammar of Japanese.

[16]The syntactic derivation is split into its subcomponents for the sake of a more readable font size. The third derivation shows how these two subparts are put together to complete the derivation of the whole sentence.

(27) Syntax:

$$\dfrac{\dfrac{\dfrac{\text{Ken ga}}{NP_n}\quad\dfrac{\text{watasi-ta}}{NP_n\backslash NP_d\backslash NP_a\backslash S_{[-N,-T]}}<}{NP_d\backslash NP_a\backslash S_{[-N,-T]}}\quad\dfrac{\dfrac{\text{no}}{(\$\backslash S_{[-N,-T]})\backslash(\$\backslash S_{[+N,-T]})}\quad\dfrac{\text{wa}}{(\$\backslash S_{[+N,-T]})\backslash(\$\backslash S_{[+N,+T]})}}{(\$\backslash S_{[-N,-T]})\backslash(\$\backslash S_{[+N,+T]})}\text{FC}}{NP_d\backslash NP_a\backslash S_{[+N,+T]}}<$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\text{sono hon o}}{NP_a}}{S_{[+T]}/(NP_a\backslash S_{[+T]})}\text{TR}\quad\dfrac{\dfrac{\text{Mari ni}}{NP_d}}{(NP_a\backslash S_{[+T]})/(NP_d\backslash(NP_a\backslash S_{[+T]}))}\text{TR}}{S_{[+T]}/(NP_d\backslash(NP_a\backslash S_{[+T]}))}\text{FC}\quad\dfrac{\text{da}}{(S_{[+T]}/Y)\backslash(Y\backslash S_{[-T]})}}{(NP_d\backslash(NP_a\backslash S_{[+T]}))\backslash S_{[-T]}}<$$

$$\dfrac{\dfrac{\text{Ken ga watasi-ta no wa}}{NP_d\backslash NP_a\backslash S_{[+N,+T]}}\quad\dfrac{\text{sono hon o Mari ni da}}{(NP_d\backslash(NP_a\backslash S_{[+T]}))\backslash S_{[-T]}}}{S_{[-T]}}<$$

Semantics:

$$\dfrac{\dfrac{\dfrac{\dfrac{\text{Ken ga}}{k}\quad\dfrac{\text{watasi-ta}}{\lambda z\lambda y\lambda x.gave'(z,y,x)}}{\lambda y\lambda x.gave'(k,y,x)}<\quad\dfrac{\dfrac{\text{no}}{\lambda f.f}\quad\dfrac{\text{wa}}{\lambda f.\theta'(f)}}{\lambda f.\theta'(f)}\text{FC}}{\theta'(\lambda y\lambda x.gave'(k,y,x))}<\quad\dfrac{\dfrac{\dfrac{\dfrac{\text{sono hon o}}{\iota x.book'(x)}}{\lambda P.P(\iota x.book'(x))}\text{TR}\quad\dfrac{\dfrac{\text{Mari ni}}{m}}{\lambda R.R(m)}\text{TR}}{\lambda S.S(m)(\iota x.book'(x))}\text{FC}\quad\dfrac{\text{da}}{\lambda f.\rho'(f)}}{\rho'(\lambda S.S(m)(\iota x.book'(x)))}<}{\rho'(\lambda S.S(m)(\iota x.book'(x)))(\theta'(\lambda y\lambda x.gave'(k,y,x)))}<$$

$$gave'(k,m,\iota x.book'(x))$$

Here, the topicalized sentence is looking for two missing arguments, unlike in the previous derivations. Those two arguments, which occur in the 'clefted' focus position, are both type-raised and then combined via FC before combining with *da*. The rest of the derivation is the same as in the single constituent cases. The combination of TR and FC does not enable too flexible a range of strange constituents; only those that are groups of missing arguments from the verb are licensed based on the resource sensitivity of the underlying logical system. For example, the following derivation shows a failed attempt of combining a cluster of an accusative NP and a dative NP with a sentence from which only an accusative NP is missing; the failure in this case is due to the type mismatch between the functor and the argument at the step indicated by *.

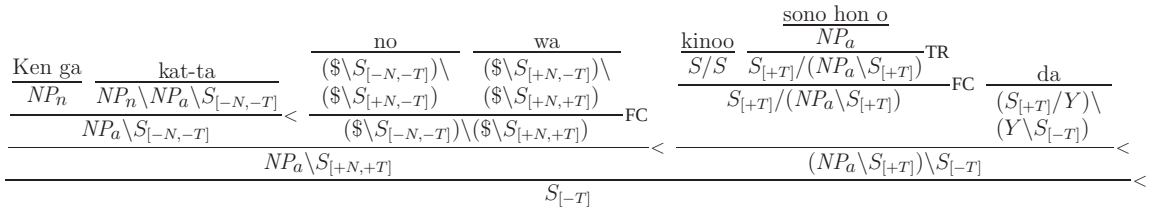
(28)

$$\dfrac{\dfrac{\text{Ken ga kat-ta no wa}}{NP_a\backslash S_{[+N,+T]}}\quad\dfrac{\text{sono hon o Mari ni da}}{(NP_d\backslash(NP_a\backslash S_{[+T]}))\backslash S_{[-T]}}}{S_{[-T]}}*$$

Further attempts to create a match prove just as fruitless as the one in (28).

Finally, cases of argument/adjunct mixed (non)constituents yield to the same procedure. The analysis for sentence (11) is given in (29). The semantic derivation is omitted due to space limitations but is easily reconstructable given the one-to-one correspondence between syntax and semantics.

(29) Syntax:



The only difference here from the previous multiple argument case is that the adjunct is already of the right type to function compose with the argument (which is type-raised, as in the former case).

4.4. Overgeneration

The previous two subsections have exemplified how a simple analysis within categorial grammar is capable of licensing all cases of Japanese clefting. The drawback that goes hand in hand with such a flexible system, however, is overgeneration. Some of the ungrammatical sentences given in section 2 are capable of being derived in this framework, but only those in which the would-be clefted constituent is on the left periphery of the corresponding main clause. Thus, example (12) (given again here as (30)) is ruled out in SOV but not OSV order.

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

A completely satisfactory general solution to the problem would be one that introduces multi-modal technology into the grammar, as in the Multi-Modal CCG work of Baldridge (2002). If there were two modes of combination (i.e. a second slash type), and one were nonassociative, it would be possible to control the way in which words can combine with one another by fine-tuning the lexicon with the newly introduced nonassociative mode. Further work is required to determine how plausible this additional mechanism is within the grammar of Japanese.

5. Conclusion

The flexibility in constituency that the cleft construction in Japanese exhibits is somewhat surprising from a purely syntactic perspective. From the point of view of the (cross-linguistically universal) function of the cleft construction in terms of its information structural difference from its simple sentence counterpart, however, the flexibility that is admitted is not really surprising. How to deal with these surprising constituents has eluded the three modern analyses of Japanese clefting presented in section 3. Section 4 presented a unified treatment of Japanese clefting wherein only three basic combinators of categorial grammar and information provided in the lexicon were used. Derivations for three general types of clefts found in the data illustrated the analysis (other cases, such as single or multiple adjunct clefts, are equally well accounted for without any additional assumptions but

were omitted due to space limitations). Thus, a framework allowing flexible constituents is shown to yield a highly explanatory solution to the problems of Japanese clefting, including an incorporation of its information structural properties.

References

Baldridge, Jason. 2002. *Lexically Specified Derivational Control in Combinatory Categorial Grammar*. Ph.D. thesis, University of Edinburgh.

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*, 153–198. Dordrecht: Kluwer.

Fukui, Naoki and Hiromu Sakai. 2003. The visibility guideline for functional categories: verb raising in Japanese and related issues. *Lingua* 113:321–375.

Fukushima, Kazuhiko. 2003. Verb-raising and numeral classifiers in Japanese: Incompatible bedfellows. *Journal of East Asian Linguistics* 12:317–347.

Koizumi, Masatoshi. 1999. *Phrase Structure in Minimalist Syntax*. Tokyo: Hituzi Syobo.

Koizumi, Masatoshi. 2000. String vacuous overt verb raising. *Journal of East Asian Linguistics* 9:227–285.

Kubota, Yusuke and E. Allyn Smith. 2005. A categorial grammar analysis of Japanese clefts: Flexibility in constituency. MS. The Ohio State University.

Martin, Samuel E. 1975/1983. *A Reference Grammar of Japanese*. New Haven and London: Yale University Press.

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

Steedman, Mark. 2000a. Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31(4):649–689.

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

Takano, Yuji. 2002. Surprising constituents. *Journal of East Asian Linguistics* 11:243–301.

The Ohio State University
Department of Linguistics
222 Oxley Hall
1712 Neil Avenue
Columbus, OH 43210-1298

kubota@ling.osu.edu, esmith@ling.osu.edu