

Analyzing Japanese Cleft Construction in Combinatory Categorical Grammar[†]

Kohei Kajikawa
The University of Tokyo
kohei-kajikawa@g.ecc.u-tokyo.ac.jp

Introduction In the Japanese cleft construction, multiple noun phrases (NPs) can occupy the focus position even if they do not form a “constituent” in the mainstream generative grammar. However, a single NP with the nominative case marker *ga* cannot.¹ Kubota and Smith (2006, 2007) analyze the cleft construction with Combinatory Categorical Grammar (CCG, Steedman, 1996, 2000; Baldrige, 2002), but their analysis overgenerates a *ga*-marked NP in the focus position. Indeed, they recognized the obligatory omission of the nominative case marker in that position, and they assumed that some independently motivated principles should explain the distribution. However, it would be better if the distribution could be explained within the grammar formalism. This study aims to address this issue by partially incorporating the idea of *constructivist* analysis of argument structure from the mainstream generative grammar (Kratzer, 1996) into the CCG framework. Furthermore, I will show that this revision correctly predicts two syntactic phenomena where the *ga*-marked NP behaves differently from other case marked NPs.

Background The Japanese cleft construction has the general form *X no wa Y da*. In this structure, *X* is a topicalized sentence with a gap, while *Y* is a focus phrase that functions as its filler. (1-a) has a single NP *sono hon-o* ‘that book-ACC’ in the focus position, while (1-b) has multiple NPs *Mary-ni sono hon-o* ‘Mary-DAT that book-ACC’.

- (1) a. Ken-ga Mary-ni watasi-ta no wa sono hon-o da.
Ken-NOM Mary-DAT give-PAST NMLZ TOP that book-ACC COP
‘It was that book that Ken gave to Mary.’
b. Ken-ga watasi-ta no wa Mary-ni sono hon-o da.
Ken-NOM give-PAST NMLZ TOP Mary-DAT that book-ACC COP
‘It was that book to Mary that Ken gave.’

Kubota and Smith (2006, 2007) proposed a CCG-based analysis of the Japanese cleft construction which can derive not only a single-focus cleft but also a multiple-foci cleft. As an example, the derivation of the single-focus cleft according to the analysis of Kubota and Smith (2006) is illustrated below:

[†]My sincere thanks to Yusuke Kubota and Shinnosuke Isono for many helpful comments. I also thank the anonymous reviewers for their insightful comments. This work is supported by JSPS KAKENHI 18K12393 and the NINJAL collaborative research projects “Cross-linguistic Studies of Japanese Prosody and Grammar” and “Evidence-based Theoretical and Typological Linguistics”.

¹In this paper, following Hiraiwa and Ishihara (2002), constructions with case marked NPs in the focus position are termed “cleft”, while those without a case marked NP are termed “pseudo-cleft”. My primary focus lies on the former. I follow Takano (2015) for the acceptability judgment of the cleft constructions used in this paper. As for the pseudo-cleft, if we interpret *no* as a pronoun, my proposed analysis can correctly derive it with slight adjustments to the category of the relativizer, as proposed in Bekki (2010, p.247)

(2)	Ken-ga	Mary-ni	watasi-ta	no wa	sono hon-o	da
	NP_{ga}	NP_{ni}	$S[-\frac{N}{T}] \setminus NP_{ga} \setminus NP_{ni} \setminus NP_o$	$(S[+\frac{N}{T}] \setminus \$) \setminus (S[-\frac{N}{T}] \setminus \$)$	NP_o	$(S[-T] \setminus X) \setminus (S[+T] \setminus X)$
	: k	: m	: $\lambda x \lambda y \lambda z. gave'(z, y, x)$: $\lambda f.f$: $\iota x. book'(x)$: $\lambda f.f$
			Perm		$\xrightarrow{>T}$	
			$S[-\frac{N}{T}] \setminus NP_o \setminus NP_{ga} \setminus NP_{ni}$		$S[+T] / (S[+T] \setminus NP_o)$	
			: $\lambda y \lambda z \lambda x. gave'(z, y, x)$: $\lambda P.P(\iota x. book'(x))$	
			$S[-\frac{N}{T}] \setminus NP_o \setminus NP_{ga}$		$S[-T] \setminus (S[+T] \setminus NP_o)$	
			: $\lambda z \lambda x. gave'(z, m, x)$: $\lambda P.P(\iota x. book'(x))$	
			$S[-\frac{N}{T}] \setminus NP_o$			
			: $\lambda x. gave'(k, m, x)$			
			$S[+\frac{N}{T}] \setminus NP_o : \lambda x. gave'(k, m, x)$			
						$S[-T] : gave'(k, m, \iota x. book'(x))$

The features $\pm N$ and $\pm T$ indicate whether the category has been nominalized and whether the category has been topicalized, respectively. In addition to the standard Function Application (“>”, “<”), Function Composition (“**B**”), and Type Raising (“**T**”) used in the usual CCG framework, Kubota and Smith also employ a permutation rule (“Perm”) as a combinatory rule to account for scrambling. The category of *no wa*, $(S[+\frac{N}{T}] \setminus \$) \setminus (S[-\frac{N}{T}] \setminus \$)$ topicalizes a sentence with a gap, and the category of *da*, $(S[-T] \setminus X) \setminus (S[+T] \setminus X)$ associates the filler with the gap in the topicalized sentence.

However, in Kubota and Smith’s analysis, they treat cases as being assigned within the lexicon. In other words, since the analysis distinguishes cases solely based on the feature-values of the *NP* category, it fails to capture the asymmetrical distribution of cases. As a result, the analysis incorrectly licenses ungrammatical sentences with a single *ga*-marked nominative NP in the focus position like (3).

- (3) *Sono hon-o Mary-ni watasi-ta no wa Ken-ga da.
 That book-ACC Mary-DAT give-PAST NMLZ TOP Ken-NOM COP
 ‘It was Ken that gave that book to Mary.’

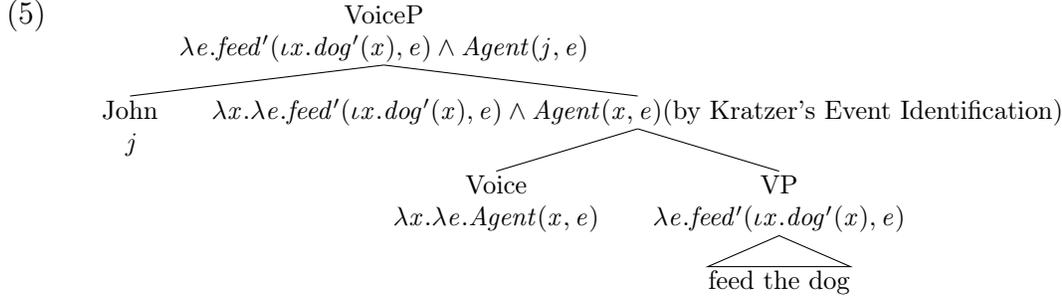
It is important to clarify that *ga*-marked NPs are not always excluded from the focus position. When the focus comprises multiple NPs, the inclusion of a *ga*-marked NP is acceptable as shown in (4). Particularly in (4-b), a *ga*-marked NP is placed immediately before the copula. Thus, it does not seem appropriate to attribute the ungrammaticality of the *ga*-marked NP in the focus position of the cleft construction to morpho-phonological constraints that ban the linear sequence of *ga da*, as argued in Takano (2015).

- (4) a. Sono hon-o watasi-ta no wa Ken-ga Mary-ni da.
 That book-ACC give-PAST NMLZ TOP Ken-NOM Mary-DAT COP
 ‘It was Ken to Mary that gave that book.’
 b. Sono hon-o watasi-ta no wa Mary-ni Ken-ga da.
 That book-ACC give-PAST NMLZ TOP Mary-DAT Ken-NOM COP
 ‘It was Ken to Mary that gave that book.’

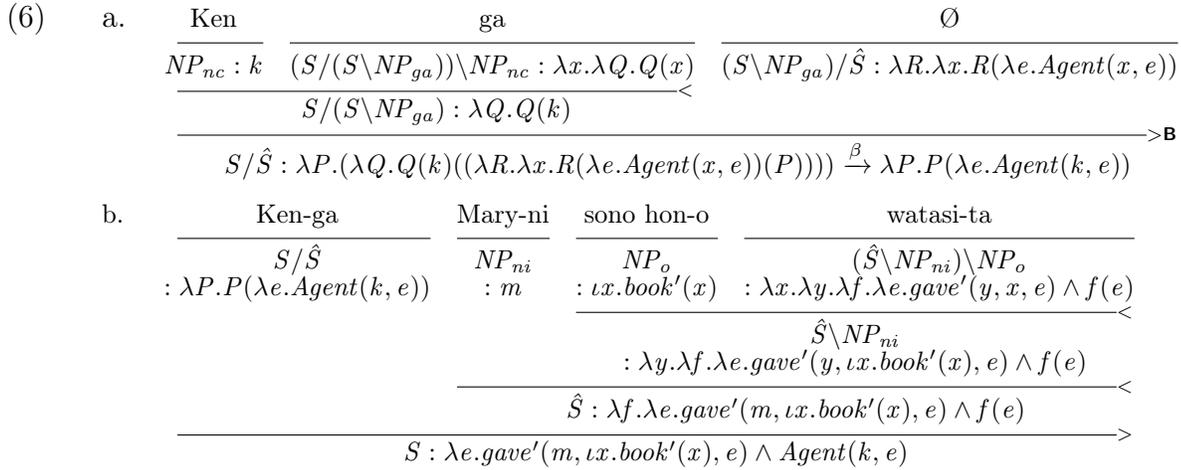
In the following section, I propose an analysis that syntactically excludes *ga*-marked NPs from the focus position of the single-focus cleft.

Proposal To treat the *ga*-case distinctly from other cases, I incorporate an idea proposed within the mainstream generative grammar into CCG in line with Isono et al. (2023). I assume that *ga*-marked NPs occupy a *structurally distinct position* compared to other case marked NPs. Within the mainstream generative grammar, Kratzer (1996) strips the verb of its ability to take an external argument and introduces a phonologically null “Voice” head as in (5). The Voice head takes a verb

phrase as its complement and introduces the external argument as its specifier. The argument structure is analyzed using the neo-Davidsonian semantics.

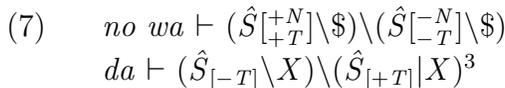


I adopt Kratzer's proposal of the Voice head to CCG as shown in (6). Regarding the semantics of the verb, I adopt the approach of Champollion (2015), ensuring that the verb binds the event variable.²



NP_{nc} represents an NP without a case marker. \hat{S} indicates the entire verb phrase without the external argument (i.e., VP in (5)). \emptyset functions similarly to Kratzer's Voice head. It becomes S when it takes \hat{S} on its right and NP_{ga} on its left. The syntactic category of ga , $(S/(S \setminus \text{NP}_{ga})) \setminus \text{NP}_{nc}$, takes NP_{nc} on its left, yielding $S/(S \setminus \text{NP}_{ga})$, which indicates that the ga -marked NP occupies a structurally higher position than other case marked NPs. Since the external argument is introduced by \emptyset , independent of the verb, the category of the verb $watasi-ta$ is $(\hat{S} \setminus \text{NP}_{ni}) \setminus \text{NP}_o$, which does not require the ga -marked NP.

The categories for *no wa* and *da* can now be revised using the \hat{S} category as shown below:



The revised CCG analysis can correctly derive both the single-focus cleft (8) and the multiple-foci cleft with the ga -marked NP (9).

²I am grateful to an anonymous reviewer for pointing out that the event variable bounded at sentence level is problematic for some compositional semantics and this can be solved by the idea of Champollion (2015).

³Another difference from the category assumed by Kubota and Smith is the direction of the slash. The vertical slash “|” signifies that either “/” or “\” can be used, and I replaced “/” in the category of *da* with this vertical slash.

(8)	Ken-ga S/\hat{S} : $\lambda P.P(\lambda e. Agent(k, e))$	Mary-ni watasi-ta $\hat{S}\backslash NP_o$: $\lambda x.\lambda f.\lambda e.gave'(m, x, e) \wedge f(e)$	no wa $(\hat{S}^{[+N]}\backslash \$)\backslash$ $(\hat{S}^{[-N]}\backslash \$) : \lambda f.f$	sono hon-o NP_o : $\iota x.book'(x)$	da $(\hat{S}^{[-T]}\backslash X)\backslash$ $(\hat{S}^{[+T]} X) : \lambda f.f$
		$\hat{S}^{[+N]}\backslash NP_o$: $\lambda x.\lambda P.\lambda e.gave'(m, x, e) \wedge P(e)$	$\hat{S}^{[+T]}/(\hat{S}^{[+T]}\backslash NP_o)$: $\lambda P.P(\iota x.book'(x))$		
			$\hat{S}^{[-T]}\backslash (\hat{S}^{[+T]}\backslash NP_o)$: $\lambda P.P(\iota x.book'(x))$		
		$\hat{S}^{[-T]} : \lambda P.\lambda e.gave'(m, \iota x.book'(x), e) \wedge P(e)$			
	$S : \lambda e.gave'(m, \iota x.book'(x), e) \wedge Agent(k, e)$				
(9)	Mary-ni watasita $\hat{S}\backslash NP_o$	no wa $(\hat{S}^{[+N]}\backslash \$)\backslash (\hat{S}^{[-N]}\backslash \$)$	Ken-ga S/\hat{S}	sono hon-o $\hat{S}^{[-T]}/(\hat{S}^{[-T]}\backslash NP_o)$	da $(\hat{S}^{[-T]}\backslash X)\backslash (\hat{S}^{[+T]} X)$
		$\hat{S}^{[+N]}\backslash NP_o$	$S/(\hat{S}^{[-T]}\backslash NP_o)$		
			$S\backslash (\hat{S}^{[+T]} NP_o)$		
	S				

Furthermore, as demonstrated in (10), the scrambling rule proposed by Bekki (2010) contributes to syntactically deriving the construction where scrambling occurs within the same clause.⁴ On the other hand, as shown in (11), the revised analysis correctly fails to derive the cleft construction when only a single *ga*-marked NP is focused. This accurately captures the distribution of the *ga*-case in the Japanese cleft construction.

(10)	Mary-ni watasita $\hat{S}\backslash NP_o$	no wa $(\hat{S}^{[+N]}\backslash \$)\backslash (\hat{S}^{[-N]}\backslash \$)$	sono hon-o NP_o	Ken-ga S/\hat{S}	da $(\hat{S}^{[-T]}\backslash X)\backslash (\hat{S}^{[+T]} X)$
		$\hat{S}^{[+N]}\backslash NP_o$	$\hat{S}^{[-T]}/(\hat{S}^{[-T]}\backslash NP_o)$		
			$(S/(\hat{S}^{[-T]}\backslash NP_o))/ (S/\hat{S}^{[-T]})$		
			$S/(\hat{S}^{[-T]}\backslash NP_o)$		
			$S\backslash (\hat{S}^{[+T]} NP_o)$		
	S				
(11)	Sono hon-o Mary-ni watasita \hat{S}	no wa $(\hat{S}^{[+N]}\backslash \$)\backslash (\hat{S}^{[-N]}\backslash \$)$	Ken-ga S/\hat{S}	da $(\hat{S}^{[-T]}\backslash X)\backslash (\hat{S}^{[+T]} X)$	
		$\hat{S}^{[+N]}$			

As the *ga*-marked NP is categorized as S/\hat{S} , it exhibits behavior distinct from NPs of other cases. S/\hat{S} can take *da* on its right side when combined with other type-raised NPs (9–10). However, when standing alone, it cannot be combined with *da* as shown in (11).

Extensions I have proposed a revised CCG analysis in which the *ga*-marked NP occupies a structurally higher position than other case marked NPs, achieved by assigning the category S/\hat{S} to the *ga*-marked NP. Furthermore, the revised analysis correctly predicts two other syntactic phenomena where the *ga*-marked NP behaves differently from other case marked NPs.

⁴Due to space constraints, I do not show in detail, but even if the scrambling occurs within a topicalized sentence like *Mary-ni Ken-ga watasi-ta no wa sono hon-o da*, it can be derived by using both Bekki's scrambling rule and a permutation rule.

First, the revised analysis predicts that the *ga*-marked NP cannot move across a clause boundary in *long-distance scrambling*. Long-distance scrambling is a construction in which an argument is scrambled out of a clause. On the assumption that Type Raising only applies to basic categories, the *ga*-marked NP cannot move out of a clause, while other case marked NPs can move out of their boundaries. Indeed, as shown in Saito (1985), the subject NPs are not subject to “long-distance scrambling” in Japanese, while other case marked NPs can be scrambled out.

Second, the revised analysis predicts that the *ga*-marked NP cannot be situated within a small clause. On the assumption that a small clause does not constitute an *S* node, then the *ga*-marked NP, which forms an *S* node in the revised analysis, cannot be placed within a small clause. This is also the case, as demonstrated by Takezawa (1987).

- (12) John-wa [Mary-no yokogao-{*ga/o} totemo utukusiku] omot-ta.
 John-TOP Mary-GEN profile-NOM-ACC very beautiful think-PAST
 ‘John thought [Mary’s profile (to be) very beautiful].’ (from Takezawa (1987, p.153))

Conclusion In this paper, I have proposed a novel analysis of the distribution of the *ga*-marked NP in the Japanese cleft construction by integrating Kratzer’s constructivist approach into CCG. As a result, I demonstrated that it is possible to account for the distribution of the *ga*-marked NP in the cleft construction within the grammar formalism. Furthermore, I showed that this analysis correctly predicts two other syntactic phenomena where only the *ga*-marked NP behaves differently from other case marked NPs. These results suggest, in line with Isono et al. (2023), that it is effective to incorporate the constructivist analysis, which has been developed within the mainstream generative grammar, into CCG.

References [1] Baldrige, Jason. 2002. Lexically specified derivational control in Combinatory Categorical Grammar. Doctoral Dissertation, University of Edinburgh. [2] Bekki, Daisuke. 2010. *Nihongo bunpoo no keisiki riron [Formal theory of Japanese grammar]*. Tokyo: Kurosio. [3] Champollion, Lucas. 2015. The interaction of compositional semantics and event semantics. *Linguistics and Philosophy* 38:31–66. [4] Hiraiwa, Ken, and Shinichiro Ishihara. 2002. Missing links: Cleft, sluicing, and “no da” construction in Japanese. *MIT Working Papers in Linguistics* 43:35–54. [5] Isono, Shinnosuke, Takuya Hasegawa, Kohei Kajikawa, Koichi Kono, Shiho Nakamura, and Yohei Oseki. 2023. Formalizing argument structures with Combinatory Categorical Grammar. In *Logic and Engineering of Natural Language Semantics*, ed. Daisuke Bekki, Koji Mineshima, and Eric McCready, 83–99. Cham: Springer Nature Switzerland. [6] Kratzer, Angelika. 1996. Severing the external argument from its verb. In *Phrase structure and the lexicon*, ed. Johan Rooryck and Laurie Zaring, 109–137. Dordrecht: Kluwer. [7] Kubota, Yusuke, and E Allyn Smith. 2006. Syntax and semantics of Japanese nonconstituent clefting in Combinatory Categorical Grammar. In *Proceedings of the Thirty-Sixth Annual Meeting of the North East Linguistic Society*, ed. C. Davis, A. R. Deal, and Y. Zabbal, 413–426. Amherst, MA: GLSA Publications. [8] Kubota, Yusuke, and E Allyn Smith. 2007. A Multi-Modal Combinatory Categorical Grammar analysis of Japanese nonconstituent clefting. In *Proceedings of the Twelfth ESSLLI Student Session*, ed. Ville Nurmi and Dmitry Sustretov, 161–172. [9] Saito, Mamoru. 1985. Some asymmetries in Japanese and their theoretical implications. Doctoral Dissertation, MIT. [10] Steedman, Mark. 1996. *Surface structure and interpretation*. MIT press. [11] Steedman, Mark. 2000. *The syntactic process*. MIT press. [12] Takano, Yuji. 2015. Surprising constituent unlabeled syntactic objects. *Nanzan Linguistics* 10:55–73. [13] Takezawa, Koichi. 1987. A configurational approach to case-marking in Japanese. Doctoral Dissertation, University of Washington.