Focus Clitics and Discourse Information Spreading

Ryo Otoguro
University of Essex

Proceedings of the LFG03 Conference

University at Albany, State University of New York

Miriam Butt and Tracy Holloway King (Editors)

2003

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In some languages, morphology plays a crucial role to represent a sentence discourse structure. In this paper, Japanese focus clitics and their distribution are examined. Against recent works on information structure, an independent discourse structure is postulated as a part of the grammar of LFG. Andrews and Manning's (1999) information spreading architecture is adopted here, so that flexible sharing of various types of information among phrase structure nodes is allowed. Moreover, Stump's (2001) Paradigm Function Morphology (PFM) functions as a parallel correspondence between phrase structure, functional structure and discourse structure. Finally, some implications to the puzzling behaviours of the similar focus clitics in Hindi are skeched.

1 Introduction

Many of the recent works on discourse function of a language claims that the grammar has an independent component representing the information structure of a sentence (e.g. Lambrecht 1994, Vallduví 1992).[1] Apart from the terminological variations (topic-comment, focus-background etc.), those works show cases where prosodic, morphological and syntactic structures (and sometimes complex of them) reflect the information structure of a sentence such as intonation, morphological discourse markers and cleft sentences. Thus, it is a natural consequence that some attempts have been made to study the relationship between the information structure and previously assumed linguistic structures like phonology and syntactic configurations as well as formalising the information structure itself. For example, Engdahl and Vallduví (1996) try to incorporate Vallduví's (1992) information packaging as a part of the framework of Head-driven Phrase Structure Grammar (HPSG: Pollard and Sag 1994). Role and Reference Grammar (RRG: Van Valin and LaPolla 1997) proposes focus structure based on the works like Lambrecht (1994).

Lexical Functional Grammar (LFG: Kaplan and Bresnan 1982, Bresnan 2001), like RRG, assumes parallel structures of the grammar. However, in the standard assumption of the framework, TOPIC and FOCUS are called discourse functions and placed in f(unctional)-structure alongside grammatical functions such as SUBJect and OBJect. As Bresnan (2001:97) notes, those TOPIC and FOCUS are not a part of discourse in a sense of communicative functions like information packaging, but syntactically represented (grammaticalised) functions. Some works try to pursue how those grammaticalised discourse functions capture configurationally distributed discourse structures. Based on the careful examinations of the configurational representations of topic-focus information in Hindi/Urdu and Russian, King (1997) and Butt and King (to appear) propose an independent structure for discourse functions instead of analysing them in the discourse functions in f-structure. Choi (1999) indicates the possibility of postulating the discourse structure as a part of the LFG grammar.

Following those preceding works, I also assume d(iscourse)-structure as a part of the parallel LFG grammar. Based on this assumption, I focus on the relationship between morphology and d-structure in this paper. Although the interface between morphology and discourse information has been less studied in LFG, Sharma (2003) analyses interesting behaviours of Hindi focus clitics adopting Nordlinger's (1998) constructive morphology. I present an analysis of the similar phenomena in Japanese focus clitics based on a different approach to morphology, namely Stump's (2001) Paradigm Function Morphology (PFM). PFM is one of the realisation models of morphology where a bundle of morphosyntactic features

[1]I would like to express my gratitude to Andrew Spencer for the comments and discussions from the earlier version of this paper. I also thank the audience in the LFG03 conference, particularly Joan Bresnan, George Aaron Broadwell, Miriam Butt, K.P. Mohanan, Tara Mohanan and Rachel Nordlinger. Of course, remaining errors are mine.

receives formal realisations by the morphological component. I place PFM as a part of the LFG architecture following the works such as Luís et al. (2002) and Sadler and Nordlinger (to appear). Under the current proposal, the morphology component is accessible to f-structure and d-structure, takes features of each structure as an input and spells out the output form in c-structure.[2] This proposal shows the striking contrast to the standard LFG assumption of morphology where a lexical items is inserted into syntax as a fully inflected form. That is, morphological operations are pre-syntactic in the standard LFG, while the c-structure configuration, f-structure and d-structure are visible to morphology in the current proposal. I will show that this model explains morphological manifestations of discourse information neatly.

The paper is structured as follows. I briefly overview the information structure summarised in Lambrecht (1994) and recast it in LFG grammar in section 2. In section 3, the general description of the morphological markings of discourse function is introduced, and the previous approaches to the clitics are examined. The morphological analysis of Japanese data and its implications to Hindi data are presented in section 4. I conclude the discussion in section 5.

2 What is discourse structure?

Lambrecht (1994:5) defines the discourse (information) structure as "the component of sentence grammar in which propositions as conceptual representations of state of affairs are paired with lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as unit of information in given discourse contexts." The crucial point of this definition is that the discourse structure is tied to lexicogrammatical structure, namely it must be linguistically observable. In other words, the discourse structure is reflected in prosodic and morphosyntactic structures. This leads to the justification of postulating the discourse structure as an independent component of the grammar rather than an extra-linguistic structure. This point is stated as "just as there are no sentences without morphosyntax and phonological structure, there are no sentences without information structure" (Lambrecht 1994:16).

The information structure is constructed by adding or superimposing the assertion on the presupposition. Those two concepts are defined as follows (Lambrecht 1994:52):[3]

(1) a. Pragmatic presupposition: The set of propositions lexicogrammatically evoked in a sentence which the speaker assumes the hearer already knows or is ready to take for granted at the time the sentence is uttered.

b. Pragmatic assertion: The proposition expressed by a sentence which the hearer is expected to know or take for granted as a result of hearing the sentence uttered.

As in (1), the presupposition is one or another formally evoked by the speaker in the sentence, namely lexicogrammatically evoked. So it must receive formal manifestations, such as prosodic prominence, morphological marking and a particular syntactic position.

[2] As discussed in Sadler and Spencer (2001) and Sadler and Nordlinger (to appear), we need a distinction between syntactic (grammatical) features and morphological (formal) features. However, I do not try to formalise the mapping between those two types of features. Rather, I use trivial mapping between them in most of the cases.

[3] As Lambrecht notes, presupposition corresponds to 'old information' and assertion to 'new information'. However, those two concepts are about the proposition as a whole and must not equated with the lexical or phrasal elements out of which propositions are formed. In other words, we cannot say a particular NP or VP is old/new information ([±new]) since the old/new distinction of each element is irrelevant (or at least not directly related) to constructing the information structure. See Lambrecht (1994:45-50) for the detailed discussion on this point.

TOPIC and FOCUS are information structure categories indicating relations between referents and propositions.[4] The TOPIC and FOCUS structure of the proposition where the referent is an argument determines the correlation between the grammatical (prosodic, morphological and syntactic) structure and discourse referents. Thus, I postulate this TOPIC and FOCUS structure as a part of the LFG grammar called d(iscourse)-structure. The definitions of TOPIC, topic expression (topic phrase, topic constituent), FOCUS and focus domain are as follows (Lambrecht 1994:131, 213, 214):

(2) a. TOPIC: A referent is interpreted as the topic of a proposition if in a given situation the proposition is construed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee's knowledge of this referent.

b. Topic expression: A constituent is a topic expression if the proposition expressed by the clause with which it is associated is pragmatically construed as being about the referent of this constituent.

c. FOCUS: The semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition.

d. Focus domain: A syntactic domain in a sentence which expresses the focus component of the pragmatically structured proposition.

Under the current study, we are concerned about how d-structure information interacts with other components of grammar. Thus, as for formalisation in LFG, I assume that the d-structure contains the semantic (LCS) information of the topic and focus expressions.[5] Technically, functional schemata ($\downarrow$LCS) $\in$ ($\uparrow$ $_{\delta}$TOPIC) and ($\downarrow$LCS) $\in$ ($\uparrow$ $_{\delta}$FOCUS) assigned to the topic/focus constituents abstract LCS values of the domain and map them onto the d-structure (cf. King's (1997) PRED FN abstraction).

As simple illustrations, information structures are exemplified in the following three way distinctions:

(3) a. What happened to your car?
My car/It [$_{VP}$ broke DOWN].

b.

Sentence:	*My car broke DOWN*.
Presupposition:	"speaker's car is a topic of comment x"
Assertion:	"x = broke down"
Focus:	"broke down"
Focus domain:	VP

(4) a. I heard your motorcycle broke down.
My [$_{NP}$ CAR] broke down.

b.

Sentence:	*My CAR broke down*.
Presupposition:	"speaker's x is broke down"
Assertion:	"x = car"
Focus:	"car"
Focus domain:	NP

[4]The cognitive states of referents themselves are defined by different information structure categories ACTIVATION and IDENTIFIABILITY (Lambrecht 1994:109). It is possible to include those states in d-structure, but I leave this possibility open.

[5]The standard LFG notion TOPIC in f-structure corresponds to Lambrecht's TOP (left-detached topic constituents). Lambrecht also proposes A-TOP ("Antitopic") for right-detached topic constituents. I do not discuss those types of topic expressions in this paper.

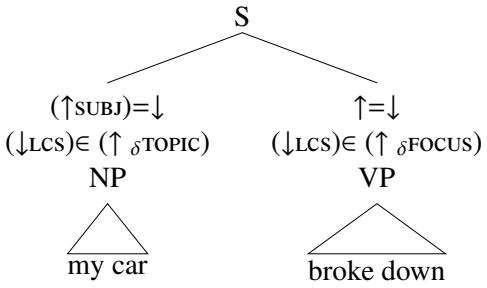
(5) a. What happened?
[$_S$ My CAR broke down].

b.

Sentence:	*My CAR broke down*.
Presupposition:	–
Assertion:	"speaker's car broke down"
Focus:	"speaker's car broke down"
Focus domain:	S

Those are called predicate focus (3), argument focus (4) and sentence focus (5) respectively. Let us look at how predicate focus example (3) can be represented in the current proposal.

(6) a. c-structure



b. f-structure

$$\begin{bmatrix} \text{LCS} & \left[\text{Event BREAK DOWN } (X)\right] \\ \text{TERMS} & \langle X \rangle \\ \text{SUBJ} & \left[\begin{matrix} \text{LCS} & \left[\text{Object MY CAR}\right] \\ \text{TERMS} & \langle\ \rangle \end{matrix}\right]^X \\ \text{TENSE} & \text{PAST} \end{bmatrix}$$

c. d-structure

$$\begin{bmatrix} \text{FOCUS} & \left\{\left[\text{LCS} \quad \left[\text{Event BREAK DOWN } (X)\right]\right]\right\} \\ \text{TOPIC} & \left\{\left[\text{LCS} \quad \left[\text{Object MY CAR}\right]\right]\right\} \end{bmatrix}$$

In the mapping to d-structure, the NP *my car* is annotated as TOPIC, so the LCS value of this constituent is mapped onto TOPIC. Similarly, LCS value of the VP *broke down* is mapped onto FOCUS in the d-structure.

3 Marking the discourse information

In English examples of the previous section, the morphology and syntax of three types of focus structure are identical. Instead, English realises the information structure by intonation in many cases. However, the strategies of information structure manifestation differ among languages. In this section, I briefly look at how languages mark discourse information morphologically. (7) and (8) are Navajo and Turkish examples respectively (Schauber 1978: 148, 152, König 1991: 17):

(7) a. Jáan chidí**ísh** yiyíiɬchǫ
John car.QFOC 3.3.PAST.wreck
*'Did John wreck the car?'
'Is it the car that John wrecked?'

b. Jáan chidí yiyíiɬchǫ'**ísh**
John car 3.3.PAST.wreck.QFOC
'Is it wrecking John did to the car?'

c. Jáan **hanii** chidí yiyíiɬchǫ'
John NEGFOC car 3.3.PAST.wreck
'It's not John who wrecked the car.'
*'John didn't wreck the car.'

(8) a. oraya ben **de** gittim
there I FOC go.PAST
'I too went there.'

b. ben oraya **da** gittim
I there FOC go.PAST
'I went there too (as well as elsewhere).'

c. ben oraya gittim **de**
I there go.PAST FOC
'I also went there.'

Navajo has question focus marker *ísh* and negative focus marker *hanii*. As seen in (7a, b), *ísh* indicates what element in the sentence is questioned, namely 'car' in (7a) and the wrecking event in (7b). In the same way, *hanii* should follow the element negated as in (7c). Turkish also has focus marker *d-*, which specifies what is focused in the sentence discourse structure as found in (8).

Hindi and Japanese also have morphological ways of realising information structure by attaching particles to a particular constituent. I overview them in the next section.

3.1 The basic description of particles

Hindi and Japanese have a set of markers, which is traditionally called 'particles'. Those particles are attached to the host nominal and represent certain grammatical properties.[6] They are mainly divided into case particles and discourse particles. The basic usage of those particles is found in (9) and (10):

(9) Hindi

a. Mōmbattī=**to** milī, lēkin ab^{h} māchis gum gayē.
the candle.F.NOM-TOP found.PERF.F.SG but now match.NOM lost go-PERF.PL
'The candle was found but now the matches are lost.'

b. Rād^{h}ā=nē=**hī** bacc^{h}õ=kō kahdānī sunāyī.
Radhaa-ERG-EXCL FOC children-ACC story.F make hear.PEF.F.SG
'It was (only) Radha who told the children a story' (Sharma 2003:61, 62)

(10) Japanese

a. Taroo=**wa** Hanako=ni yubiwa=o ageta.
Taroo-TOP Hanako-DAT ring-ACC give.PST
'Taro gave Hanako a ring.'

[6]Some of the particles can appear with other categorical hosts, such as verb, adjectives and even clauses. In this paper, however, I mainly focus on the attachment to nominals.

b. Ken=**mo** titioya=o=**sae** nikunda.
Ken-FOC father-ACC-FOC hate.PST
'Ken too hated even his father.'

Particles are attached to the nominal hosts in each example as indicated by =. The noun with particles carries certain information in the sentence. For instance, *mōmbattī=tō* in (9a) and *Taroo=wa* in (10a) are topic in the information structure of each sentence. *Titioya=o=sae* in (10b) is a direct object and focus at the same time. Similarly, *Rād^{h}ā=nē=hī* in (9b) functions as a subject and focus. The following is a part of the basic set of case and discourse particles in those two languages (cf. Sharma 2003, Butt and King in press):

(11)

Hindi	Japanese	
o	*ga*	nominative
nē	–	ergative
kō	*o*	accusative
k-	*no*	genitive
kō	*ni*	dative
se	*de*	instrumental
hī	*dake*	exclusive contrastive focus ('only')
b^{h}ī	*mo*	inclusive contrastive focus ('also', additive/scalar)
tō	*wa*	contrastive topic
tak	*sae*	scalar endpoint marker ('even')
	made	
b^{h}ar		entirety ('all')

The phenomena analysed in this paper is focus clitic attachments of each language. I restrict the data to the following two sets of examples in this paper (% indicates that the acceptance of this sentence is subject to dialectal variation.).

(12) a. in tīn laḍkõ=kō=**hī** chōṭ lagī
these three boys=DAT=FOC hurt-F be-applied-to-PERF.F.SG
'(Only) *these three boys* got hurt.'

b. (%) in tīn laḍkõ=**hī**=kō chōṭ lagī
'(Only) these three *boys* got hurt.'

c. in tīn=**hī** laḍkõ=kō chōṭ lagī
'(Only) these *three* boys got hurt.'

d. in**hī** tīn laḍkõ=kō chōṭ lagī
'(Only) *these* three boys got hurt.' (Sharma 2003:67)

(13) a. kotira=no san'nin=no syoonen=ni=dake kega-sase-ta
these=GEN three=GEN boy=DAT=FOC hurt-CAUS-PAST
'Only *these three boys* got hurt.'

b. kotira=no san'nin=no syoonen=dake=ni kega-sase-ta
'Only *these three boys* got hurt.'

c. kotira=no san'nin=dake=no syoonen=ni kega-sase-ta
'Only these *three* boys got hurt.'

d. ?kotira=dake=no san'nin=no syoonen=ni kega-sase-ta
'Only *these* three boys got hurt.'

According to Sharma, in all the examples in (12), the focused constituent is the whole noun phrase 'these three boys' whereas the semantic scope of 'only' covers the italicised elements. In other words, there is a mismatch of the focus/semantic scope. Unlike Hindi examples, the order between dative case marker *ni* and focus marker *dake* does not seem to change the scope of 'only' in Japanese examples (9a, b), namely 'only' takes scope over the whole noun phrase in both examples.[7]

3.2 Phrase structural status

The particles summarised above are attached to the right of the host. There are three possibilities of the phrase structural status of them: X^0-level suffix, postposition and phrasal-level clitic element. Many preceding works suggest that they are clitics (Mohanan 1994, Butt and King in press, Sharma 2003, Ohara 2000), and some works in derivational frameworks assume that semantic particles are postpositions occupying P nodes in the phrase structure (e.g. Miyagawa 1989). The arguments against X^0-level suffix is found in co-ordination.

(14) a. Yasin=nē [kutt-ē or g^hoṛ-ē] =kō/hī hɛ.
Yassin.M.SG-ERG dog-M.SG.OBL and horse-M.SG.OBL ACC/FOC be.PRES.3.SG
'Yassin saw (only) the dog and the horse.'

b. Nadya [lahor or karac^{h}i] =se hɛ.
Nadya.F.SG.NOM Lahore and Karachi INST be.PRES.3.SG
'Nadya is from Lahore and Karachi.'

(15) a. Taroo=wa [inu=to uma] =o mita.
Taroo-TOP dog-and horse ACC see.PST
'Taro saw the dog and the horse.'

b. Ken=wa [suugaku=to buturi] =de/sae manten=o totta.
Ken-TOP math-and physics INST/FOCUS full mark-ACC get.PST
'Ken got full marks (even) for math and physics.'

(14) is taken from Butt and King (in press) and Sharma (2003). (14a) suggests that accusative case particle *kō* and focus particle *hī* take the scopes over the co-ordinated NPs, which is impossible for X^0-level suffixes. Similarly, semantic case *se* can have the scope over the co-ordinated NPs as in (14b). Japanese example (15) shows exactly the same behaviour. Another argument against suffix status is accentuation. For example, the lexical stress on each NP in Japanese is not affected by attachment of the particles. This lack of stress interaction is observed in Hindi as well. According to those behaviours of the particles, the possibility of X^0-level suffix is rejected.[8]

[7] The order can affects the interpretation of the sentences with the relation to the predicates (Harada and Noguchi 1992). The detailed semantic analysis of those examples should be treated in semantic structure and are beyond the scope of this paper. So, I simplify the picture here and regard (9a, b) as free variations.

[8] Hindi incorporated focus markers in (12d) are inseparable from the host. I will come back to this point later.

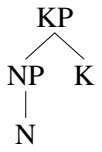
The possibility of independent words, i.e. postposition, is also refuted. Although the particles in Hindi and Japanese have phrasal scope in co-ordinated structure, they cannot stand alone without the hosts, that is they are bound elements. So, for example, scrambling or wh-fronting of a noun phrase leaving the particle behind is impossible. In addition, Hindi has postpositions which may stand themselves. (Butt and King in press).

The conclusion drawn from the preceding observations is that the particles in Hindi and Japanese are clitics. They are bound element, but phrasal attachment is possible (Zwicky 1987, Zwicky and Pullum 1983). The next question to ask is whether clitics occupy syntactic terminals in phrase structure or not, and if so how they project.

3.3 Projection of clitics and constructive morphology

Butt and King (in press) propose that a case clitic is a syntactic object occupying a terminal node K(ase) and becomes a head of KP taking NP complement, i.e. host noun phrase, in Hindi/Urdu as in (16):

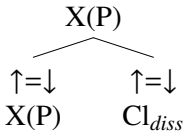
(16)



Butt and King claim that K can contribute complex of features associated with case, including grammatical function and semantically relevant material such as volitionality.

Sharma (2003) follows Butt and King's assumption for case clitics. As for discourse clitics, she proposes the structure like (17).

(17)

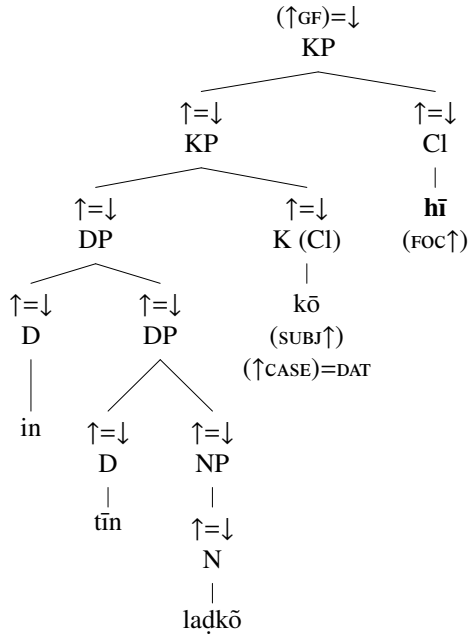


(18) a. *nē* (SUBJ↑)
(↑CASE)=ERG

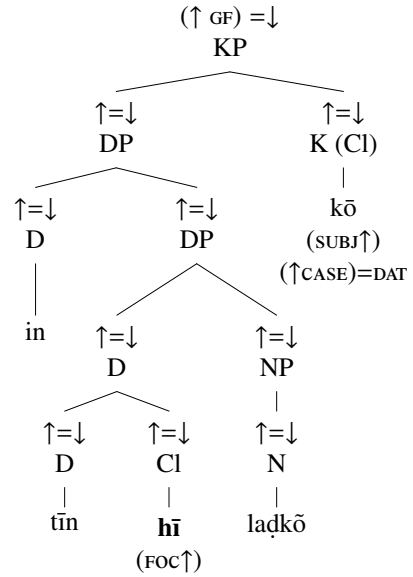
b. *hī* (FOC↑)

Sharma assumes that Hindi focus clitics can be attached to X^0-level, so that Cl is adjoined to either X or XP. The host and clitic are co-heads, namely LFG notation ↑=↓ is attached to the sister nodes XP and Cl. Further, she expands constructive morphology (Nordlinger 1998) to focus clitics, so (FOC↑) is associated with the lexical entry *hī* as in (18b) in the same way as (SUBJ↑) in the ergative case marker (18a). (18b) states that the f-structure containing *hī* functions as FOCUS in the outer clause. Based on those assumptions, Sharma analyses Hindi focus clitic examples (12) like (19).

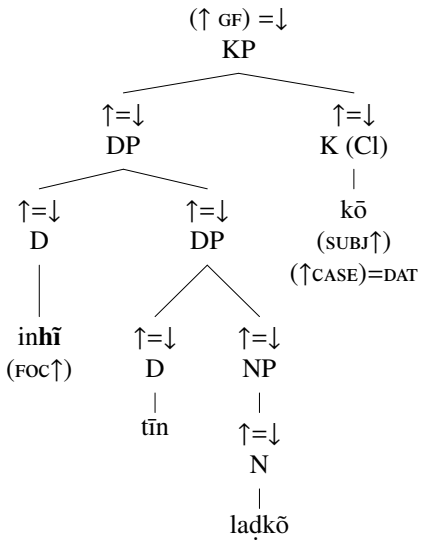
(19) a.



b.



c.



d.

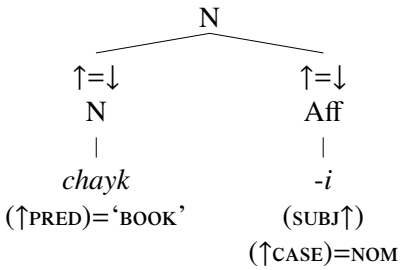
$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{CASE} & \text{DAT} \\ \text{PERS} & 3 \\ \text{NUM} & \text{PL} \\ \text{COUNT} & 3 \\ \text{DEIX} & \text{PROX} \\ \text{PRED} & \text{`boys'} \end{bmatrix} \\ \text{FOC} & [\ ] \\ \text{PRED} & \text{`}\ldots\langle\ldots\rangle\text{'} \end{bmatrix}$$

However, there is a difficulty in the KP projection analysis proposed by Butt and King. C-structure in LFG is surface phrase structure – place-holder of words. Therefore, even though the K node is associated with complex feature bundle, it does not justify introducing functional projection like KP. Such an assumption causes a further problem. Since Hindi and Japanese have other types of clitics like discourse and quantification clitics, the functional projection analysis would assume that we could have other types of projection such as FocusP, TopicP, ConjP and QP headed by clitics. This means c-structure contains many functional information, which normally belongs to other structures like f-structure or discourse-structure. Even if we do not assume such functional projections for other types of clitics like Sharma, it is mystery why only case clitics can be a head of KP projection. Or if K covers the other types of clitics, it would be still unclear what the status of K is and why it functions as a head of a functional projection.

Sharma's analysis has other difficulties. Firstly, she assumes that the prenominal modifiers are D and a head of the functional projection DP. It takes either an NP complement or another DP complement and holds a co-head relation. This co-head relation and functional DP projection are crucial for Sharma's analysis to pass (FOC↑) to the top KP node, so that the whole KP is focused. However, as argued in Fukui (1986, 1995), Börjars (1999) and many others, the status of prenominal modifiers are not straightforwardly determined. At least, the difference of the clitic attachment and the incorporation found in (19b, c) indicates the possibility that they constitute distinct classes. Thus, it is skeptical that they are really a head of the functional DP projection.

As Andrews and Manning (1999) point out, the co-headness of the standard LFG is sometimes problematic, and Sharma acknowledges a problem of the mismatch of focus scope and semantic scope. Although we can apply more specific information sharing of Andrews and Manning's (1999) model, a deep problem seems to lie in the morpheme-based assumption of LFG. LFG is based on what Stump (2001) calls 'incremental-lexical' model. Lexical items including morphemes are stored in the lexicon and paired with a particular information. Morphemes are combined with the hosts below X^0 (or X^0-level). The information carried by each item is projected into f-structure by function ϕ. So, for instance, Lee (1999)[9] proposes that Korean case marker is adjoined to the host N and the N and this case marker become co-head:

(20)



Those types of analyses are equally applicable to the treatment of inflectional morphology in general in the standard LFG. So, for example, Bresnan (2001:57) shows the f-description of the lexical entry *lion*, *live*, plural noun suffix *-s* and verb agreement suffix *-s* as follows:

(21) a. *lion*: N (↑PRED) = 'LION'
-s: $infl_N$ (↑NUM) = PL

b. *live*: V (↑PRED) = 'LIVE ⟨. . .⟩'
-s: $infl_V$ (↑TENSE) = PRES
(↑SUBJ) = ↓
(↓PERS) = 3
(↓NUM) = SG

What (21) suggests is plural marker *-s* carries the information indicating its mother, i.e. N, is plural whereas verbal suffix *-s* carries the information specifying its mother's (V) tense is present, and the person and number values of the mother's subject is 3rd and singular respectively.

The works in realisational model of morphology (Matthews 1972, 1991, Anderson 1992, Aronoff 1994, Beard 1995, Stump 2001) have pointed out the problems found in the assumption of one-one pair

[9]See also Sells (1995), Cho and Sells (1995), Andrews (1996), Nordlinger (1998) and many others for a similar treatment of morphological operations in LFG.

between form and function as in (20, 21) in that it is hard to capture some morphological phenomena such as cumulation, multiple exponence and so on.[10] Contrary to incremental-lexical approaches, the basic assumption behind 'inferential-realisational' model of morphology is that it regards the formatives, such as English *-s* in (21), as formal realisation of morphosyntactic features. In other words, the realisational model rejects the idea of treating morphemes as Saussurean signs where one-one pair between form and function is assumed. Thus, the inflections like (21) are not concatenation of base verb *like* and person/number agreement morpheme *-s*, or base noun *lion* and plural morpheme *-s* by the rule of morphotactics. Rather, lexeme LIKE and LION inflect for the morphosyntactic properties associated with them such as person, number and tense, and the morphology changes the forms of the lexemes to realise those properties, in (21) by suffixation to the root.[11]

Against those realisational model background, the solution to problematic Hindi and Japanese clitics is straightforward, that is they are inflectional suffixes at the phrasal level. We can assume that they do not occupy syntactic terminal such as K, Cl, Prt, rather they are products of suffixation process at phrasal level by morphology.[12] The apparent difference between Hindi and Japanese nominal clitics and English suffixes in (21) is their levels of realisation, i.e. X^0 and XP. The extension of affixasation to phrasal level is favoured in many works in the line of realisational morphology (e.g. Anderson 1992). In fact, the assumption that the morphology accesses to the phrasal level and realises the morphosyntactic properties by adding morphophonological objects like clitics in Hindi and Japanese gives unified account to the data, as we will see in the following sections. I show how to attain those phrasal inflection adopting Stump's (2001) Paradigm Function Morphology (PFM) along the line found in Spencer (2000, 2003b) and Luís and Spencer (to appear).

4 The proposals

4.1 The input to syntax and phrase structure

To incorporate PFM as a morphological component of LFG and realise phrasal inflection, a modifications of the frameworks is required. While the input to syntax is thought to be a fully inflected form of the word in the standard LFG (e.g. Bresnan 2001: 44), under the current proposal it is a lexeme. I assume that the lexicon is a storage of lexemes, where each has morphological information (phonological form and permissible morphological features), syntactic information (TERMS list or argument structure, X-bar category etc.) and semantic information (LCS). Thus, morphological operations are carried out on the c-structure according to the values of morphological features, that is they are not pre-syntactic. This process is illustrated in detail in the following sections.

The combinations of lexemes are constrained by phrase structure rules as normally assumed in LFG. For the phrase structure of Japanese nominals, I follow the proposals by Fukui (1986, 1995) which is also adopted by Sells (1995) for Japanese, Cho and Sells (1995) for Korean and Nordlinger (1998) for Wambaya phrase structures. One of the crucial points in Fukui's proposals is that empirical data suggests Japanese lacks D category and accordingly lacks functional projection DP. Lacking functional category D means the projection never reaches to the XP (=X″)-level and lacks Specifier. Thus, Japanese nominal projection is open in the sense that it leaves the projection as N′. Based on this proposal of

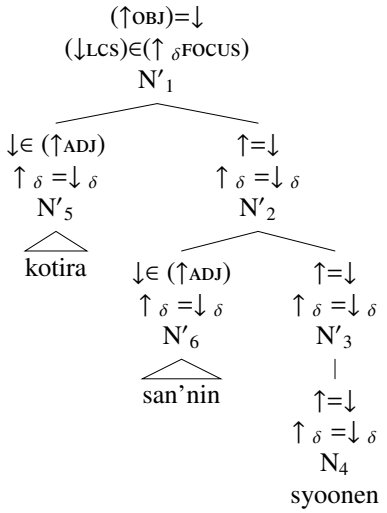
[10] For LFG, Spencer (2003a) points out the difficulties of morpheme-based lexicalism.

[11] Capital letters are used for a index of a lexeme.

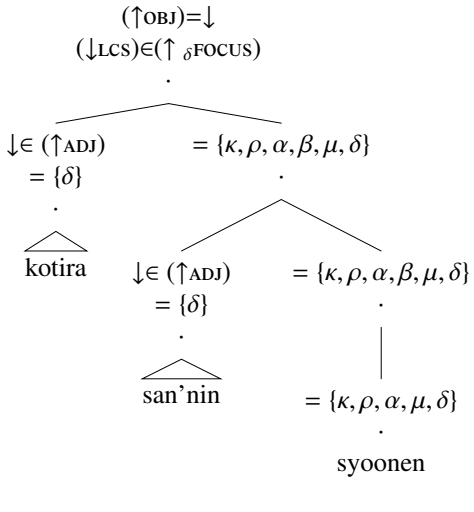
[12] We could introduce syntactic terminal by morphology as sketched in Stump (1997). But I do not take this option here.

Japanese phrase structure, the c-structure for (13) is like (22a) – or (22b), the version fully committed to information spreading.[13]

(22) a.



b.



The prenominal modifiers such as KOTIRA 'this/these' and SAN'NIN 'three' are not D. They are adjoined to N′ of the head noun SYOONEN 'boy' and annotated as ADJUNCTIONS. Note that case/focus clitics do not appear in (22) since the terminal elements are not inflected yet. They go into the morphological component and receive particular forms according to the features associated with them as explained in the next section. The corresponding f- and d-structures are as follows:

(23) a.

$$\begin{bmatrix} \text{OBJ} & \begin{bmatrix} \text{LCS} & [\text{Object BOY}] \\ \text{TERMS} & \langle\ \rangle \\ \text{NUM} & \text{PL} \\ \text{ADJ} & \left\{ \begin{bmatrix} \text{LCS} & [\text{Property THESE}] \end{bmatrix}, \begin{bmatrix} \text{LCS} & [\text{Property THREE}] \end{bmatrix} \right\} \end{bmatrix} \\ \ldots \end{bmatrix}$$

b.

$$\begin{bmatrix} \text{FOCUS} & \left\{ \begin{bmatrix} \text{LCS} & [\text{Object BOYS}] \end{bmatrix}, \begin{bmatrix} \text{LCS} & [\text{Property THESE}] \end{bmatrix}, \begin{bmatrix} \text{LCS} & [\text{Property THREE}] \end{bmatrix} \right\} \\ \text{TOPIC} & \{\ldots\} \end{bmatrix}$$

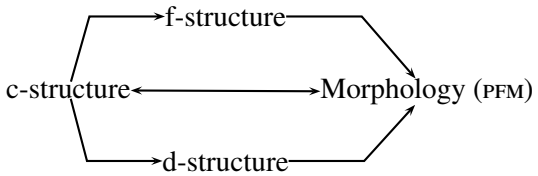
4.2 Incorporating PFM into LFG

One of the roles of morphology is giving a proper form to the stripped-off c-structure introduced in the previous section. Roughly, the general picture of the current proposal is schematised as in (24).[14]

[13] I add δ for discourse information to the natural classes assumed in Andrews and Manning (1999). The other classes are: κ (X-bar categories), ρ (grammatical functions), α (argument structure, TERMS, LCS), β (bar-level), μ (morphosyntactic features).

[14] Only the relevant components appear here. More components such as prosodic structure and semantic structure would come into the picture as well.

(24)



The morphology (PFM) part takes a lexeme (or complex of lexemes) in c-structure and the associated morphosyntactic/ discourse features as an input, then spells out a proper form for that input, which is returned to the c-structure. Note that this operation must not be regarded as derivational process, rather it is a well-formedness condition defined by PFM.

The core assumption of PFM is that function called 'Paradigm Function (PF)' specifies the inflected form for classes of lexeme of the language and places it in the correct cell in the paradigm. In Spencer's (2003b) revised version of PFM, PF is defined by a set of functions such as the rules of exponence, the rules of referral, stem selection and placement. So the basic representation of PF is like (25):[15]

(25) PF($\langle$£, $\sigma\rangle$)=$_{def}$
- i. stem: MOR($\langle$£, $\sigma\rangle$)
- ii. exponence: $R_I(\sigma)$
- iii. placement: align(R_I, Right/Left, Max/Min, stem(£, σ))

Here PF is defined by the stem selection function 'MOR', realisation rule 'R' and alignment function 'align' for the input pair $\langle$£, $\sigma\rangle$ where £ is an index of the lexeme and σ is a complete set of morphosyntactic properties associated with the lexeme.

As for Japanese nominal inflection under the current discussion, the stem selection is straightforward, i.e. a root form. A realisation rule takes a bundle of morphosyntactic features as an argument and gives the output exponent for a particular inflectional class. Some of the realisation rules for Japanese nominals are like (26):

(26)
- a. $R_{I,\{[Case:Dat]\},N}(\sigma)$ = ni
- b. $R_{I,\{[Dis:Foc],[Lcs:Only]\},N}(\sigma)$ = dake
- c. $R_{II,\{[Case:Gen]\},N}(\sigma)$ = no

The realisation rule 'R' is defined by morphosyntactic properties such as [Case], [Number], [Lcs] and [Dis(course)] it formally realises. The subscript indices such as I and II specify the rule blocks which realisation rules belong to. Each rule block contains a number of realisation rules and which rule is applied is determined by Pāṇini's principle, that is for the input feature bundle σ the most narrowly specified rule is applied. Two or more realisation rules can be combined and turn out to be a composed function. For example, if the features associated with the input lexeme are [Case: Gen], [Dis: Foc] and [Lcs: Only], (26b) and (26c) are picked up from the rule blocks I and II respectively, then those two rules become a composed function which gives a morphophonological object *dakeno* as in (27):

(27) $R_I \circ R_{II}(\sigma)$ = dakeno

Alignment function specifies where the formatives realisation rules spell-out are placed for the input lexeme. (28) is an example for the Japanese clitic attachment – for the input pair $\langle$£, {[Case: Dat]}$\rangle$.

[15] I add a hierarchical parameter (Max/Min) to the proposal of Spencer (2003b).

(28) align(R_I, Right, Max, stem(£, {[Case: Dat]}))

Function 'align' consists of 4-tuple. R is a function of the realisation rules, Right is a value of the horizontal parameter (Right/Left) determining to which direction of the host the output form of the realisation rule is attached, Max is a value of the vertical parameter (Max/Min) specifying the projection-level of the host lexeme where the output form is attached, and finally stem(£, σ) is a target stem of the lexeme. Thus, in this case 'align' states that 'place the output form of R_I – *ni* by (26a), to the right of the maximal projection of the stem of the input lexeme £'.

4.3 The analysis

We have looked at the basic mechanism of the morphology component in the previous section. I will show how this mechanism works for the Japanese nominal clitics (13), which is repeated here as (29).

(29) a. kotira=no san'nin=no syoonen=ni=dake kega-sase-ta
these=GEN three=GEN boy=DAT=FOC hurt-CAUS-PAST
'Only *these three boys* got hurt.'

b. kotira=no san'nin=no syoonen=dake=ni kega-sase-ta
'Only *these three boys* got hurt.'

c. kotira=no san'nin=dake=no syoonen=ni kega-sase-ta
'Only these *three* boys got hurt.'

d. ?kotira=dake=no san'nin=no syoonen=ni kega-sase-ta
'Only *these* three boys got hurt.'

The base c-structure for those examples are (22). [Case: Dat] is associated with the projection of the head noun SYOONEN.[16] The projections of the prenominal modifiers KOTIRA and SAN'NIN have [Case: Gen]. Since the focus covers the whole noun phrase, we assume that δ spreads among all the nodes, i.e. [Dis: Foc] is shared among the head noun projection and the prenominal modifier projections. As indicated by the italics in (29), [Lcs: Only] is included in different lexemes in the examples. In (29a, b), SYOONEN and its projection have [Lcs: Only], whereas KOTIRA and SAN'NIN and their projections have this feature in (29c) and (29d) respectively.

Let us look at how each lexeme goes into the morphology and PF puts the input onto the proper cells in the paradigm, i.e. how PF makes it well-formed according to the associated properties. (30) shows the process that PFM specifies the inflected forms of (29a, b).

(30) a. PF(⟨SYOONEN, {[Dat], [Foc], [Only]}⟩)=$_{def}$
i. stem: MOR(⟨SYOONEN, {[Dat], [Foc]}⟩)
ii. exponence: ($R_{I,\{[Dat]\},N} \circ R_{I,\{[Foc],[Only]\},N}$)({[Dat], [Foc], [Only]})
iii. placement: align($R_I \circ R_I$, Right, Max, stem(SYOONEN, {[Dat], [Foc], [Only]}))

b. PF(⟨KOTIRA, {[Gen]}⟩)=$_{def}$
i. stem: MOR(⟨KOTIRA, {[Gen]}⟩)
ii. exponence: $R_{II,\{[Gen]\},N}$({[Gen]})
iii. placement: align(R_{II}, Right, Max, stem(KOTIRA, {[Gen]}))

[16] It is more plausible to say that case is a syntactic/ grammatical feature for the noun phrase rather than the noun. Here, [Case] should be regarded as a morphological/ formal feature associated with a lexeme.

c. PF(⟨SAN'NIN, {[Gen]}⟩)=$_{def}$
   i. stem: MOR(⟨SAN'NIN, {[Gen]}⟩)
   ii. exponence: $R_{II,\{[Gen]\},N}(\{[Gen]\})$
   iii. placement: align(R_{II}, Right, Max, stem(SAN'NIN, {[Gen]}))

In (30aii), the composed function specifies the exponent for the input features like '$(R_I \circ R_I)(\sigma)$=dakeni/nidake' by (26a, b).[17] The alignment function (30aiii) shows where this exponent is placed. Since it says the output form is placed at the right of the maximal projection of the stem of the input lexeme, *dakeni/nidake* is placed at N'_1 in (22a). Similarly, according to '$R_{II,\{[Gen]\},N}(\sigma)$=no' in (26c) and placement of (30b, c), *no* is added to N'_5 and N'_6 of (22a). Thus, the resultant inflected form is [*kotira=no san'nin=no syoonen*]=*ni=dake/dake=ni*.

In the case where 'only' takes its scope over the prenominal modifiers, KOTIRA for instance, the following process gives the correctly inflected forms (29c).

(31) a. PF(⟨SYOONEN, {[Dat]}⟩)=$_{def}$
   i. stem: MOR(⟨SYOONEN, {[Dat]}⟩)
   ii. exponence: $R_{I,\{[Dat]\},N}(\{[Dat]\})$
   iii. placement: align(R_I, Right, Max, stem(SYOONEN, {[Dat]}))
   b. PF(⟨KOTIRA, {[Gen], [Foc], [Only]}⟩)=$_{def}$
   i. stem: MOR(⟨KOTIRA, {[Gen], [Foc], [Only]}⟩)
   ii. exponence: $(R_{I,\{[Foc],[Only]\},N} \circ R_{II,\{[Gen]\},N})(\{[Gen], [Foc], [Only]\})$
   iii. placement: align($R_I \circ R_{II}$, Right, Max, stem(KOTIRA, {[Gen], [Foc], [Only]}))
   c. PF(⟨SAN'NIN, {[Gen]}⟩)=$_{def}$
   i. stem: MOR(⟨SAN'NIN, {[Gen]}⟩)
   ii. exponence: $R_{II,\{[Gen]\},N}(\{[Gen]\})$
   iii. placement: align(R_{II}, Right, Max, stem(SAN'NIN, {[Gen]}))

Here, (31a) specifies *ni* is attached to N'_1 and (31b) states that *dakeno* is placed at N'_5. Those morphological operations give correct result (29c). (29d) where 'only' takes its scope over 'three' is inflected in a similar way.

4.4 Implications to Hindi focus clitics

The proposal in the previous section is applicable to Hindi data (12) without a significant modification. However, one of the intriguing aspects of the Hindi focus clitic is incorporation. As found in (12d), if *hī* is attached to the demonstrative *in*, it is incorporated. As a result, it is not separable from the host and receives the phonological effect. In other words, the focus marker is no longer a clitic. Sharma (2003:65) illustrate a set of incorporated forms of personal and demonstrative pronominals, a part of which is shown in (32):

[17] Unlike the standard PFM, I allow realisation rules in one rule block to be applied recursively (cf. Otoguro (2003) for Japanese nominal recursive rule blocks).

(32)

UNFOCUSED FORM	FOCUSED FORM	GLOSS OF FOCUSED FORM
muj^{h}	*mujhi*	me-FOC (obl.)
tum/tuj^{h}	*tumhi/tujhi*	you-FOC (obl.)
yah	*yahi*	he/she/it-FOC (prox.)
is	*isi*	he/she/it-FOC (obl., prox.)
ham	*hamhĩ*	I/we-FOC
in	*inhĩ*	they-FOC (obl., prox.)

Although I do not present a detailed analysis of Hindi incorporation, the data present some indication of the continuity between clitic and incorporated form. The fact that the incorporation is restricted to the closed class of hosts, i.e. personal/demonstrative pronominals, suggests that the morphological component (placement function, for example) is sensitive to the inflectional classes and gives formal realisations of morphosyntactic properties differently – in this case, it places *hī* onto pronominals at non-phrasal level, which triggers incorporation. This is exactly the point syntactic analyses miss as shown in the attachment of the clitic to D of (19b) and the incorporated form of (19c). In fact, such a continuity is often found in languages (e.g. Modern Greek clitics and affixes (Condoravdi and Kiparsky 2001)) and I believe that the division between clitics and affixes is not crystal clear, so it can be misleading to regard clitics as syntactic objects. Once we acknowledge that some types of formatives, which are often referred to a kind of 'clitics', are morphological objects rather than syntactic one, we can capture the generalisation and continuity by morphological operations. [18]

5 Conclusions

This paper addresses the problematic cases of a syntactic treatment of discourse clitics and their distribution in Hindi and Japanese. The alternative approach I have presented is a morphological treatment of the phenomenon. Unlike the standard LFG assumption to inflectional morphology ('incremental-lexical'), I adopt one of the 'inferential-realisational' models, Stump's (2001) PFM. Under the current proposal, the morphology is accessible to other components of grammar such as c-structure, f-structure and d-structure, so that Paradigm Function specifies the correct form of the input pair, index of the lexeme and associated features. Further, Andrews and Manning's (1999) information spreading allow flexible information sharing, i.e. semantic, discourse information, among phrase structure nodes. Such an architecture attains phrasal-level inflection and neatly accounts for the distributions of the nominals cliitics in Japanese. Further, the framework suggests the possibility of capturing the continuity of the phrasal/lexical attachment of the morphological objects, which is widely observable in languages.

Some attempts of proposing the realisation-based morphological theory within LFG have been made. A different way of presenting the morphology-syntax interface within LFG and PFM is presented by Ackerman and Stump (to appear). Sells (to appear) shows how OT-LFG framework can be combined with the realisation models of morphology. The current proposal is one of the contributions to those attempts.

References

Ackerman, Farrell, and Gregory T. Stump. to appear. Paradigms and periphrastic expression: A study in realizaion-based lexicalism. In L. Sadler and A. Spencer (Eds.), *Projecting Morphology*. Stanford,

[18] Note that I do not claim that every type of so-called clitics should be treated by morphology.

CA: CSLI.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Andrews, Avery D. 1996. Semantic case-stacking and inside-out unification. *Australian Journal of Linguistics* 16:1–54.

Andrews, Avery D., and Christopher D. Manning. 1999. *Complex Predicates and Information Spreading in LFG*. Stanford, CA: CSLI.

Aronoff, Mark. 1994. *Morphology by Itself*. Cambridge, MA: The MIT Press.

Beard, Robert. 1995. *Lexeme-Morpheme Base Morphology*. Albany, NY: SUNY Press.

Börjars, Kersti. 1999. *Feature Distribution in Swedish Noun Phrases*. Oxford: Blackwell.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Butt, Miriam, and Tracy Holloway King. in press. The status of case. In V. Dayal and A. Mahajan (Eds.), *Clause Structure in South Asian Languages*. Oxford: Oxford University Press.

Butt, Miriam, and Tracy Holloway King. to appear. Null elements in discourse structure. In K. V. Subbarao (Ed.), *Papers from the NULL Seminar*. Delhi: Motilal Banarasidas.

Cho, Young-Mee Yu, and Peter Sells. 1995. A lexical account of inflectional suffixes in Korean. *Journal of East Asian Linguistics* 4:119–174.

Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford, CA: CSLI.

Condoravdi, Cleo, and Paul Kiparsky. 2001. Clitics and clause structure. *Journal of Greek Linguistics* 2:1–39.

Engdahl, Elisabet, and Enric Vallduví. 1996. Information packaging in HPSG. In C. Grover and E. Vallduví (Eds.), *Edinburgh Working Papers in Cognitive Science*, Vol. 12, 1–31. Edinburgh: University of Edinburgh.

Fukui, Naoki. 1986. *A Theory of Category Projection and Its Applications*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Fukui, Naoki. 1995. *Theory of Projection in Syntax*. Stanford, CA: Kurosio/CSLI.

Harada, Yasunari, and Naohiko Noguchi. 1992. On the semantics and pragmatics of *dake* (and *only*). In C. Barker and D. Dowty (Eds.), *Proceedings of the SALT II*, Vol. 40 of *The Ohio State University Working Paper in Linguistics*, 125–144.

Kaplan, Ronald M., and Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*, 173–281. Cambridge, MA: The MIT Press.

King, Tracy Holloway. 1997. Focus domains and information-structure. In M. Butt and T. H. King (Eds.), *Proceedings of the LFG97 Conference*, Stanford, CA. CSLI.

König, Ekkehard. 1991. *The Meaning of Focus Particles: A Comparative Perspective*. London: Routledge.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Lee, Hanjung. 1999. Aspectual and thematic licensing of grammatical case. In *Papers from Regional Meeting of the Chicago Linguistic Society*, Vol. 35, 203–222, Chicago. Chicago Linguistic Society.

Luís, Ana, Louisa Sadler, and Andrew Spencer. 2002. Phrasal affixation and the morphology/syntax interface. Unpublished ms., presented at the LFG02, Athens, 3-5 July 2002.

Luís, Ana, and Andrew Spencer. to appear. A Paradigm Function account of 'mesoclisis' in European Portuguese. *Yearbook of Morphology*.

Matthews, Peter H. 1972. *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.

Matthews, Peter H. 1991. *Morphology*. Cambridge: Cambridge University Press. second edition edition.

Miyagawa, Shigeru. 1989. *Structure and Case Marking in Japanese*. Vol. 22 of *Syntax and Semantics*. San Diego: Academic Press.

Mohanan, Tara. 1994. *Argument Structure in Hindi*. Stanford, CA: CSLI.

Nordlinger, Rachel. 1998. *Constructive Case: Evidence from Australian Languages*. Stanford, CA: CSLI.

Ohara, Masako. 2000. *An Analysis of Verbal Nouns in Japanese*. PhD thesis, University of Essex.

Otoguro, Ryo. 2003. Nominal clitic ordering and recursive rule blocks. Unpublished ms., presented at the Linguistics Association of Great Britain Spring Meeting, University of Sheffield, 14-16 April 2003.

Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Sadler, Louisa, and Rachel Nordlinger. to appear. Relating morphology to syntax. In L. Sadler and A. Spencer (Eds.), *Projecting Morphology*. Stanford, CA: CSLI.

Sadler, Louisa, and Andrew Spencer. 2001. Syntax as an exponent of morphological features. In G. Booij and J. van Marle (Eds.), *Yearbook of Morphology 2000*, 71–96. Dordrecht: Kluwer Academic.

Schauber, Ellen. 1978. Focus and presupposition: A comparison of English intonation and Navajo particle placement. In D. J. Napoli (Ed.), *Elements of Tone, Stress and Intonation*, 144–173. Washington DC: Georgetown University Press.

Sells, Peter. 1995. Korean and Japanese morphology from a lexical perspective. *Linguistic Inquiry* 26:277–325.

Sells, Peter. to appear. Syntactic information and its morphological expression. In L. Sadler and A. Spencer (Eds.), *Projecting Morphology*. Stanford, CA: CSLI.

Sharma, Devyani. 2003. Nominal clitics and constructive morphology in Hindi. In M. Butt and T. H. King (Eds.), *Nominals: Inside and Out*, 59–84. Stanford, CA: CSLI.

Spencer, Andrew. 2000. Verbal clitics in Bulgarian: A paradigm function approach. In J. Grijzenhout and B. Gerlach (Eds.), *Clitics in Phonology, Morphology and Syntax*, 355–386. Amsterdam: John Benjamins.

Spencer, Andrew. 2003a. Paradigm-based lexicalism: Negation in Japanese. Unpublished ms., University of Essex.

Spencer, Andrew. 2003b. Revised Paradigm Function Morphology. Unpublished ms., University of Essex.

Stump, Gregory T. 1997. Template morphology and inflectional morphology. In G. Booij and J. van Marle (Eds.), *Yearbook of Morphology 1996*, 217–241. Dordrecht: Kluwer Academic.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Vallduví, Enric. 1992. *Information Component*. New York: Garland.

Van Valin, Jr., Robert D., and Randy J. LaPolla. 1997. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Zwicky, Arnold M. 1987. Suppressing the Zs. *Journal of Linguistics* 23:133–148.

Zwicky, Arnold M., and Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 59:502–513.

Ryo Otoguro
Department of Language and Linguistics
University of Essex
Wivenhoe Park
Colchester, CO4 3SQ
United Kingdom
rotogu@essex.ac.uk
http://privatewww.essex.ac.uk/~rotogu/