

Universität Bielefeld
Fakultät für Linguistik und Literaturwissenschaft

Masterarbeit
im Studiengang Linguistik:
Kommunikation, Kognition und Sprachtechnologie

Case, Word Order, Grammatical Function and Information Structure in Japanese

Vorgelegt von
Kyoko Sugisaki

Erstgutachter Prof. Dr. Ralf Vogel
Zweitgutachterin Prof. Dr. Miriam Butt

November 2010

Acknowledgements

I thank my professors and mentors Ralf Vogel and Miriam Butt. Ralf Vogel provided me constantly with support and encouragement. I benefited a lot from the discussions with him. He was the person who evoked my interest in syntax and OT and impressed me with his prompt and insightful analysis of linguistic data. Furthermore, I thank Miriam Butt for her professional comments and advice. I feel very lucky to have met Miriam Butt at the computational linguistics fall school at the University of Konstanz. She is such an inspiring teacher that I decided to write my masters thesis in the framework of LFG.

Many thanks go to my friends and fellow students at the University of Bielefeld for many good discussions. A special thank goes to Attila Sepsí and Ioan Ghisoiu for cheering me up during my master course.

I am deeply grateful to my mother and Japanese informant Nobuko Sugisaki for being patient to be tested over a dozen of Japanese sentences. My wholehearted thanks go to my husband Andreas Bleiker for being supportive and understanding during my study in Bielefeld. I would not have finished my study in two years without his support and pragmatic suggestions.

List of Abbreviations

ACC	accusative case	LOC	locative case
ADJ	adjunct	NEG	negative
AG	agent	NOM	nominative case
ASP	aspect	NZER	nominalizer
BCK	background	OBJ	object
COM	comitative case	OBJ ₀	semantically restricted object
COMP	complement	OBL	oblique
CONTRA	contrastiveness	OBL ₀	semantically restricted oblique
COP	copula	OSV	object-subject-verb
DAT	dative case	PAST	past tense
DF	discourse function	PAT	patient
EMP	emphasis	PRED	predicate
FOC	focus	PRES	present tense
GEN	genitive case	Q	question word
GER	gerund	SOV	subject-object-verb
GF	grammatical function	SUBJ	subject
HON	honorific	TOP	topic
I.OBJ	indirect object	VOL	volitional
INSTR	instrumental case	VP	verbal phrase
K	kase(case)	XADJ	open adjunct
KP	kase (case) phrase	XCOMP	open complement

Index

Acknowledgements

List of Abbreviations

1. INTRODUCTION	5
1.1. OVERVIEW	5
1.2. THEORETICAL ASSUMPTION	7
1.2.1. LEXICAL-FUNCTIONAL GRAMMAR	7
1.2.2. OPTIMALITY THEORETICAL LEXICAL-FUNCTIONAL GRAMMAR	10
2. GRAMMATICAL FUNCTIONS IN JAPANESE	12
2.1. PHRASE STRUCTURE AND GRAMMATICAL FUNCTION IN JAPANESE	12
2.2. MORPHOLOGY AND GRAMMATICAL FUNCTION	19
3. INFORMATION STRUCTURE IN JAPANESE	27
3.1. INFORMATION STRUCTURE	27
3.2. REALIZATION OF TOPIC, FOCUS AND CONTRASTIVENESS IN JAPANESE	29
3.2.1. FOCUS	29
3.2.2. TOPIC	36
3.3. INFORMATION STRUCTURE IN LFG	44
4. OT-LFG	51
4.1. INPUT	52
4.2. ARGUMENT STRUCTURE AND GRAMMATICAL FUNCTION	53
4.3. IDENTIFICATION OF GRAMMATICAL FUNCTION	55
4.4. REALIZATION OF INFORMATION STRUCTURE	71
5. CONCLUSION	80
6. REFERENCES	82

1. Introduction

1.1. Overview

This thesis aims to examine the configurationality and word order of Japanese. Japanese is known as a free word order language, which allows arguments to be scrambled relatively freely on the condition that the verb is positioned sentence-finally. This free word order is, however, not so unrestricted that any transitive sentence has two word order options, SOV and OSV. The word order becomes fixed to encode the grammatical function if the morphology does not do so unambiguously. This word order freezing phenomena (Mohan 1994) has been observed in other free word order languages such as German (Vogel 2004), Urdu (Mohan 1994, Lee 2001a), Russian (Bloom 1999) and Korean (Lee 2001a). The word order is considered to be an interaction between the encoding of grammatical and discourse function in these languages. The word order is restricted to the canonical word order to encode the grammatical function, while it is also optimized to realize the information structure (e.g., topic-comment). Furthermore, the unmarked word order is dependent on the animacy grade in German, with regard to the direct object and indirect object (Vogel and Steinbach 1998). The word order is constructed by the dimension of the grammatical function (case), semantic property (e.g., animacy, definiteness e.g., Choi (1996)), pragmatic context (information structure) and prosodic well-formedness constraints (e.g., focus position; see Buring (2000)); weak function words like pronouns in Germanic languages (see Vogel (2006)). The restriction and optimization of the word order is thoroughly discussed in other free word order languages such as German (Choi 1996, Müller 1999, Buring 2000, Müller 2002) and Korean (Choi 1996, Lee 2001a), however, there is relatively little work that observes word order from these various perspectives in Japanese. Much has been written on the restriction of word order as a theory internal discussion in generative grammar (Saito 1989, Miyagawa 1997 among others), but only a few works from different perspectives (Ishihara 2000a, 2000b, Flack 2007). Also noteworthy is that the volume of work is done on the psycholinguistics of Japanese (e.g., Kazumitsu 1983, Muraoka, Tamaoka and Yayoi 2005).

The first aim of this thesis is to examine word order and case marking as a strategy of the encoding of grammatical and discourse function in Japanese. As a concrete framework to demonstrate how the word order and case marker works to identify the

grammatical and pragmatic functions, I use Lexical-Functional Grammar (Bresnan 2001). Lexical-Functional Grammar (LFG) is a mathematically well-defined grammar with parallel architecture, i.e., without movement. According to LFG, linguistic representations coexist in parallel and are connected to each other by mathematically defined projection. The functional structure (f-structure) is the core of the grammar and represents grammatical function (e.g., subject, object) and morphosyntactic information (e.g., case, number, and gender). The actual or surface word order is represented at the constituent/categorical-structure (c-structure). The discourse function is traditionally represented at the f-structure. However, a separate representation, i.e. i-structure (King 1997), serves as an adequate representation for the more fine-grained analysis of information structure. This thesis concentrates on f-structure, c-structure, and i-structure, and does not discuss much other representations such as argument structure (a-structure), morphological structure (m-structure), or prosodic/phonologic structure (p-structure).

The second aim is to investigate the interaction of word order and morphology with regard to the realization of the grammatical and discourse functions. I assume that Optimality-Theoretic Lexical-Functional Grammar (OT-LFG) is the most suitable framework for this analysis of word order interaction with morphology, grammatical function, and discourse function. OT-LFG (Bresnan 1996) combines the procedural and competitive conception of the grammar of Optimality Theory (Prince and Smolensky 2004) and the formal elegance of LFG. The advantage of OT-LFG is to observe the different linguistic representations at the same stage and to see how relevant an individual principle is for a particular language. The formal device of LFG provides thereby a clear description of the input, candidate set, and output.

As to the organization of this thesis, Chapter 2 discusses grammatical functions, which are encoded by phrase structure (section 1) and by morphology (section 2). The subsections in Chapter 2 deal with configurationality in the language and its representation in LFG. Chapter 3 goes on to discuss the information structure of Japanese. After defining the notion of information structure (section 1), section 2 provides an analysis of its realization via the morphological and syntactic means in Japanese. Section 3.3 provides the LFG analysis and is discussed with respect to the discourse configurationality of Japanese. The last chapter is dedicated to the interaction between morphemes, word order, grammatical, and discourse functions in OT-LFG. The next section briefly provides some basic assumptions regarding LFG and OT-LFG.

1.2. Theoretical Assumption

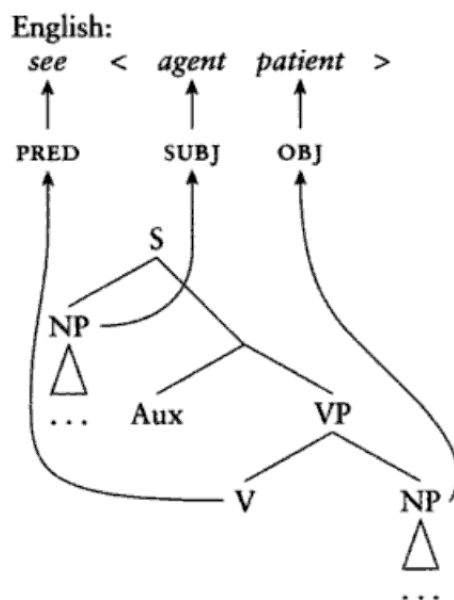
1.2.1. Lexical-Functional Grammar

This section provides the basic assumptions of LFG in brief. LFG is a formal non-derivational grammar that does not assume movement. Semantic information such as thematic role is represented at the level of argument-structure (a-structure), and the syntactic representation is divided into two structures: f(unctional)-structure and c(onstituent)-structure.

Consider the basic transitive sentence in English:

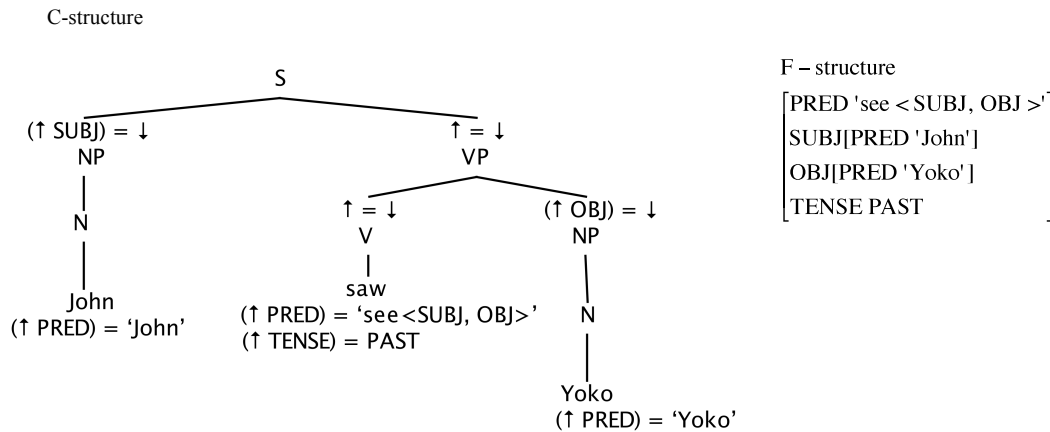
(1) John saw Yoko

(2) A-structure, f-structure, and c-structure in LFG adopted from Bresnan (2001)



The predicate *see* assigns two arguments and the semantic roles agent and patient, respectively. This layer belongs to the a-structure. The a-structure is linked to grammatical functions, which corresponds to the f-structure. The syntactic tree is the representation of the c-structure and projected from the f-structure.

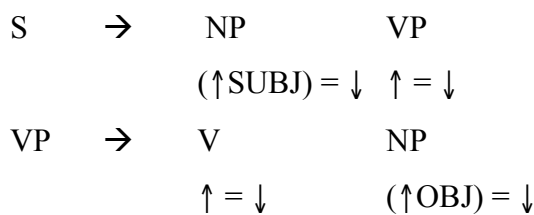
(3) 'John saw Yoko' for English in LFG



The f-structure represents morphosyntactic information such as tense and grammatical function (e.g., SUBJECT and OBJECT) and can be organized in the form of an Attribute-Value Matrix (AVM). Complementing this, the c-structure represents the actual constituent structure in tree form, whose nodes are annotated with f-annotation such as (↑SUBJ) = ↓. The f-annotation connects the c-structure to the f-structure.

LFG comprises phrase structure rules with f-annotation and lexical entries. The lexicon plays an important role in the grammar and represents the subcategorization, syntactic category, and morphosyntactic information (e.g., case, person, gender, tense, etc.).

(4) 'John saw Yoko' for English Grammar in LFG



John N (↑PRED) = 'John'
Yoko N (↑PRED) = 'Yoko'
saw V (↑PRED) = 'see <SUBJ, OBJ >'
(↑TENSE) = PAST

The phrase structure rule generates the actual word order and constituency at the c-structure. Therefore, the phrase structure does not necessarily configure the grammatical

function in the architecture of LFG. The morphological marking such as head marking or dependent marking contributes to the configuration as well.

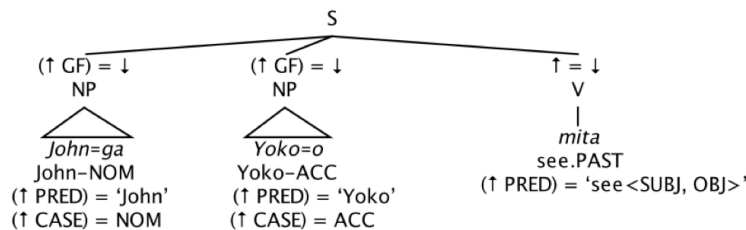
Consider the sentence (5) in Japanese:

- (5) John=ga Yoko=o mita
 John=NOM Yoko=ACC see.PAST
 ‘John saw Yoko’

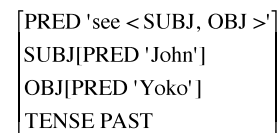
The f-structure is identical for English and Japanese. The f-structure is considered to be mainly language-independent, whereas the c-structure varies among languages (Butt et. al 1999). The c-structure is flat in Japanese.

(6) ‘John saw Yoko’ for Japanese in LFG

C-structure



F-structure



The flat c-structure is different from English in that the grammatical function is not licensed in the phrase structure position but via the case marking:

(7) ‘John saw Yoko’ for Japanese Grammar in LFG

S	→	NP*	V
		(↑GF) = ↓	↑ = ↓
John=ga		NP	(↑ PRED) = ‘John’ (↑ CASE) = NOM
Yoko=o		NP	(↑ PRED) = ‘Yoko’ (↑ CASE) = ACC
saw		V	(↑ PRED) = ‘see<SUBJ, OBJ>’ (↑ TENSE) = PAST

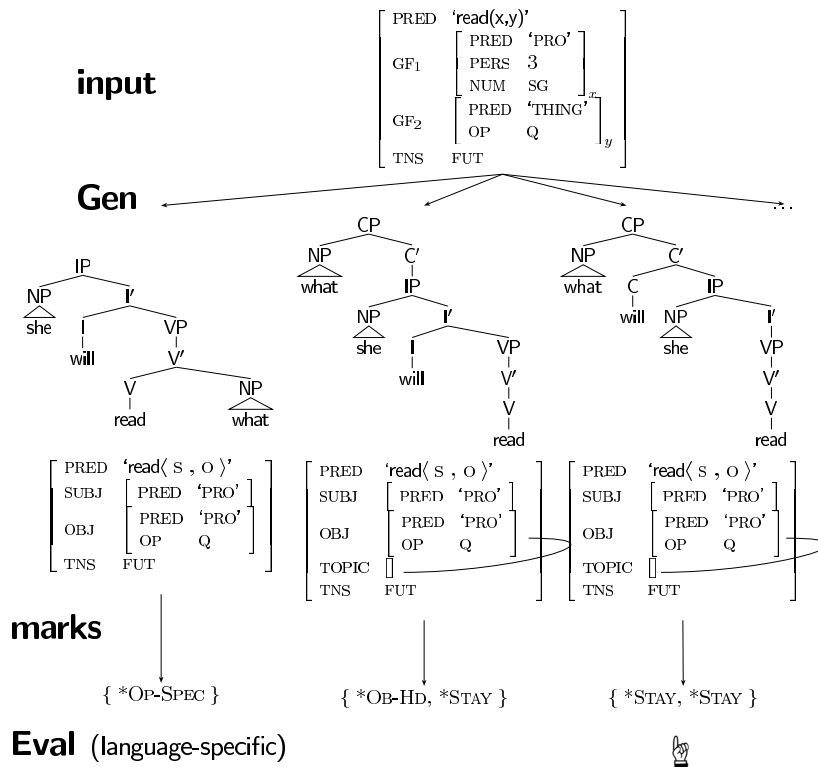
The predicate checks the case feature of the grammatical function. As for the assignment of the case, there are three possibilities in LFG: structural (or default), quirky

(or lexical), and semantic case assignment (Zaenen, Maling and Thráinsson 1985, Butt and King 2004b). Structural case assignment is the default for the semantic unrestricted grammatical function such as SUBJ or OBJ. Case can be assigned by position or by a morphological marking (Butt and King 2004b). The predicate *mita* is, for example, the case of the default assignment by using meta-rules. Quirky case is the idiosyncratic case that is assigned individually by the verb, adjective, or preposition in the lexicon (Zaenen, Maling and Thráinsson 1985). The German dative predicate *helfen* (translated: help) is, for example, the case of the quirky case. Semantic case is assigned to the semantic restricted grammatical function such as indirect object (OBJ θ) or oblique case (OBL θ).

1.2.2. Optimality Theoretical Lexical-Functional Grammar

The classical OT-LFG is production-based and generates some forms for a given meaning. As an input, it provides the argument structure and some semantic relevant information such as tense. Theoretically, the information structure is also semantically relevant and considered to be an input. GEN(erator) is the grammar which uses the input to produce some possible forms and is considered as universal and inviolable. All languages have the same candidate set (the f-structure and c-structure pair), and differences between the languages are due to the differences in the ranking of the violable constraints. Some constraint can be of high relevance for one language, but can be the irrelevant for the another language. The output is the optimal form for the given meaning which violates the fewest high-ranked constraints.

(8) Production-based OT-LFG adopted from Kuhn (2000)



The illustration in (8) shows the process of the generation *what will she read* in English. The input consists of the semantic information such as the meaning *read*, which requires two arguments. The first one is pronoun, 3rd person and single. The second one has the meaning *thing* and question word. The tense is future. The GEN generates some possible c-structure and f-structure for that meaning. The violation of the constraint OP-SPEC, OB-HD and STAY is examined and marked. The Eval(uation) decides which one is the best candidate, according to the constraint ranking for the language. Here the rightmost candidate is selected because it violates the fewest constraints that are important for the language. The constraint STAY (“don’t move”) is of less relevance for the language. The constraint OP-SPEC (“Operator in specifier”) and OB-HD (“obligatory head”) are more important than STAY. OP-SPEC lets the wh-operator *what* move into the specifier position, whereas OB-HD obligates every phrase to have an (extended) head and moves the auxiliary *will* into C.

2. Grammatical Functions in Japanese

This Chapter is dedicated to the encoding of the grammatical functions (e.g., subject, object) in Japanese in LFG. Japanese is a language that uses two ways to encode grammatical functions: the phrase structure and the morphology. This chapter discusses these strategies in the language.

Section 2.1 first examines configurationality in Japanese and sets up the basic phrase structure at c-structure. Section 2.2 provides an analysis of case marking in Japanese and its relationship to the grammatical functions.

2.1. Phrase Structure and Grammatical Function in Japanese

Since Hale (1980), there are polarized assumptions on the configurationality or non-configurationality in Japanese. According to Hale (1983), non-configurational languages have distinctive properties, i.e. (1) free word order, (2) use of syntactic discontinuous expressions, (3) extensive use of null anaphora. Accordingly, Japanese is a typical non-configurational language. The permutation of the arguments is relatively free, discontinuous constituents are highly restricted but it is possible for the quantificational expression to be separated from the noun phrase and float freely in the sentence. Finally, the argument pronouns are often omitted. Contrary to Hale's diagnostic, Saito (1985) has shown some asymmetries between subject and object as evidence for a VP node, providing the analysis on the pronominal coreference, quantifier floating, weak crossover and resumptive pronouns. Saito's argumentation for the VP node is anchored in the Government-Binding Theory, in which movement is the basic concept of the grammar and c-command plays a fundamental role in the hierarchical syntactic structure. In LFG, in contrast, the surface structure is of relevance for the c-structure, where the actual linear word order and constituency are represented.

The constituent of the VP is traditionally diagnosed using the syntactic operation such as the pro-form *so suru* (translated *do so*) in Japanese. It is, however, controversial and does not show strong evidence for the VP-node. Nakau (1973) shows that the verbal phrase can be replaced by the pro-form *soo su(ru)* as shown in the sentence (9)a. The Verb alone does not however substitute the pro-form as shown in sentence (9)b.

(9) VP-node in Japanese from Nakau (1973) cit. in Miyagawa (1989)

- a. Taroo=¹ga Telebi=o mita.
Taroo=NOM TV=ACC see.PAST
Ziro=mo soo shita.
Ziro-too so do.PAST
'Taroo watched TV and Jiro did so'
- b. *Taroo=ga Telebi=o mita.
Taroo=NOM TV=ACC see.PAST
Ziro=mo Telebi=o soo shita.
Ziro=too Telebi=ACC so do.PAST
'Taroo watched TV and Jiro did so'

Hinds (1973) however provides the counter evidence that the *soo su(ru)* can be substituted by some sequence of sentences as shown in sentence (10).

(10) Counter example from Hinds (1973) cit. in Miyagawa (1989)

- Taroo=wa kankoku=e itta; sorekara Osaka=e
Taroo=TOP Korea=LOC go.PAST and Osaka=LOC
kaetta; soshite hikooki=de Amerika=e itta;
return.PAST and airplane=INSTR America=LOC go.PAST
Hanako=mo soo shita
Hanako=too so do.PAST
'Taroo went to Korea, then returned to Osaka, and went to America by plane.
Hanako did so too'

Unlike Japanese, English shows a strong constituency of the VP. The sentence (11) shows that the object and verb build a constituency. As Fukui (1995) pointed out, Japanese allows the adverbial elements be placed between a verb and its direct object. The constituency of the object and the verb is also obvious in the pseudo-clefting in English as shown in the sentence (12). The object is fronted together with the verb but the subject is not.

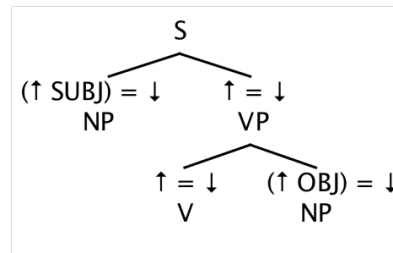
¹ I mark clitics as = in accordance with the convention. The clitic tests are provided in the next subsection.

“Configurational” and “non-configurational” refer to “languages which make primary use of syntactic strategies and morphological strategies respectively for the identification of grammatical functions” (Nordlinger, 1998:45).

Accordingly, configurational languages encode grammatical functions by the position of the phrase structure and non-configurational languages use morphological means such as the case marking or head marking. A configurational language such as English is generated by the phrase structure rule with the f-annotation like (14).

(14) Configurational Language

$S \rightarrow NP \quad VP$
 $(\uparrow \text{SUBJ}) = \downarrow \quad \uparrow = \downarrow$
 $VP \rightarrow V \quad NP$
 $\uparrow = \downarrow \quad (\uparrow \text{OBJ}) = \downarrow$

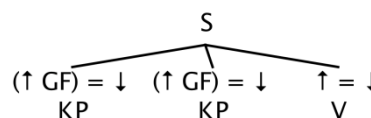


The Functional equation written in $\uparrow = \downarrow$ means that the mother’s f-structure is the same as own f-structure. $(\uparrow \text{SUBJ}) = \downarrow$ means that the NP bears the subject function and its mother’s attribute SUBJ contains the value of its node. The dominance relationship is encoded with \rightarrow . The phrase structure rule makes the specific grammatical function definite by the f-structure annotation in the certain position: the subject is an NP under S-node and the object is an NP under VP-node.

In contrast, non-configurational languages are assumed to have some morphological means to identify grammatical functions. The grammatical function is therefore underspecified as GF³ at the c-structure. For example, the phrase structure rule (15) generates a transitive sentence in Japanese, presupposes that it is a non-configurational language and has an exocentric flat structure with the exocentric category S. The category S has no lexical head and can dominate more than one argument with its predicate.

(15) Non-configurational Language

$S \rightarrow KP^* \quad V$
 $(\uparrow \text{GF}) = \downarrow \quad \uparrow = \downarrow$



³ GF can be any of the grammatical function:
 GF = {SUBJ|OBJ|OBJθ|COMP|XCOMP|OBL|ADJ|XADJ}

The specification of the grammatical function is ensured not by the position of the phrase structure but by the term, which is connected to the lexicon. For example, both sentences (16) can be generated by the phrase structure rule (15) and the lexical entries in (17). The lexical entries provide information about the f-structure such as (SUBJ↑) from the nominative case *ga*:

(16)

- a. John=*ga* Yoko=*o* *mita*
 John=NOM Yoko=ACC see.PAST
 ‘John saw Yoko’
- b. Yoko=*o* John=*ga* *mita*
 Yoko=ACC John=NOM see.PAST
 ‘John saw Yoko’

(17) Lexicon

- a. *mita*: V
 (↑PRED) = ‘see<SUBJ, OBJ>’
 (↑TENSE) = PAST
- b. John: N
 (↑PRED) = ‘John’
- c. Yoko: N
 (↑PRED) = ‘Yoko’
- d. *ga*: K
 (↑CASE) = NOM
 (SUBJ ↑)
- e. *o*: K
 (↑CASE) = ACC
 (OBJ ↑)

↑ denotes the f-structure of the mother, whose node is annotated. Hence, the annotation (↑TENSE) = PAST states for example that the f-structure of the mother has the attribute TENSE, whose value is PAST. I follow Nordlinger (1998) and assume that the case marking constructs per se the grammatical function and provides directly the

information about the grammatical function. Thereby the syntactic category K^4 for the case marking is introduced at the pre-terminal node. (SUBJ \uparrow) means that the higher f-structure corresponds to the attribute SUBJ and subsumes the actual matrix-value [CASE NOM]. That results in [SUBJ [CASE NOM]] at the f-structure. Hence, this annotation constructs the functional constraints from inside to outside and is referred to the “inside-out functional uncertainty” (Dalrymple 2001). As shown in (16), the arguments can be freely positioned in Japanese and the flat structure does not fix the position of XP with the specific grammatical function and generates both sentences (16).

The question is now what happens if the morphology does not provide information at the terminal node to encode the grammatical function. As in other free word order languages such as Hindi (Lee 2001a), Korean (Lee 2001a), or German (Choi 1996), word order freezing occurs also in Japanese (Ohara 2001, Flack 2007). The word order is fixed or freezes if the morphology is unable to determine the grammatical functions, as illustrated in (18).

- (18) Word Order Freezing in Japanese
- a. John=*ga* Yoko=*ga* *suki.da*
 John=*NOM* Yoko=*NOM* *like.COP.PRES*
 ‘John likes Yoko’
- b. ?⁵ Yoko=*ga* John=*ga* *suki.da*
 Yoko=*NOM* John=*NOM* *like.COP.PRES*
 *’John likes Yoko’
 ‘Yoko likes John’

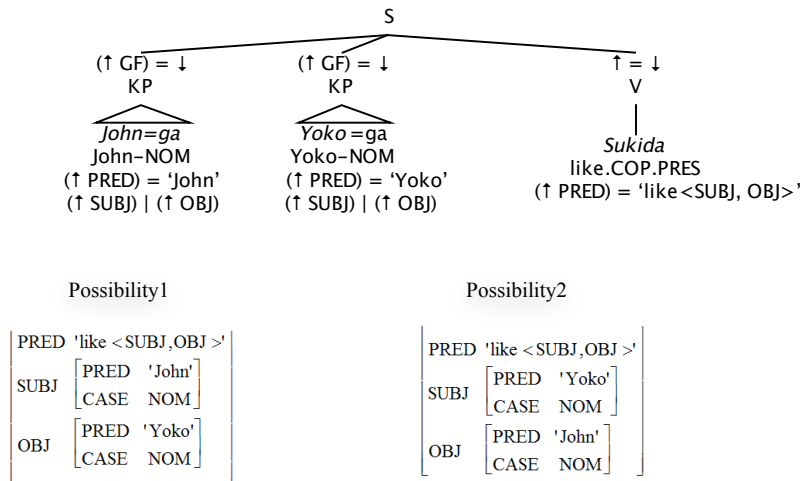
The nominative case marker *ga* marks both argument phrases, John and Yoko. Stative predicates such as the psychological predicate *like* or the ability modality *can* require the direct object to be marked by the nominative case rather than by the accusative case. The misinterpretation of the sentence (18)b shows that the freedom of word order is lost – the grammatical function is determined by syntactic means. The SOV-ordering is therefore considered as “canonical” (or “unmarked”, “neutral”) word order.

⁴ Japanese case marker is often called case particle as part of the noun phrase (see e.g., Kishimoto (2010)) or postposition as P (e.g., for example Gunji 1987). I assume here that the Japanese case marker has the functional category K. The details are discussed in the next section.

⁵ The symbol ? indicates that the sentence is grammatically correct but leads to misinterpretation or misalignment of arguments and their grammatical functions. Here, the patient Yoko is falsely interpreted as the subject and thus annotated with the question mark.

This dilemma is observed when the flat structure is suitable for the free word order language, but the morphology doesn't always function as the linker and the word order encodes the grammatical function. If the flat structure preserves and there are no structural assignments of grammatical functions, the overgeneration happens: The sentence (18)a will have two grammatical functions for John and Yoko:

(19) Problem: Word order freezing



It is necessary to use the preference of the canonical word order (SOV) to the non-canonical word order (OSV) as follows:

(20) Japanese Phrase Structure with preference⁶.

$S \rightarrow KP^* \quad V$
 $(\uparrow GF) = \downarrow \quad \uparrow = \downarrow$

S: Linear order constraint 1: SUBJ < OBJ
 CANON-ORDER $\in o::*$

Linear order constraint 2: OBJ < SUBJ
 NON-CANON-ORDER $\in o::*$

OT-Ranking: NON-CANON-ORDER CANON-ORDER

⁶ Sells (1990) proposed the following phrase structure rule in Japanese:

$S \rightarrow NP^* \quad PP^* \quad VP$
 $(\uparrow GF) = \downarrow \quad (\uparrow OBL) = \downarrow \quad \uparrow = \downarrow$
 $VP \rightarrow VP \quad AUX$
 $\uparrow = \downarrow \quad \uparrow = \downarrow$
 $VP \rightarrow NP^* \quad PP^* \quad V$
 $(\uparrow OBJ(2)) = \downarrow \quad (\uparrow OBL) = \downarrow \quad \uparrow = \downarrow$

It is crucial that the phrase structure allows the direct object to be under S or under VP. I didn't discuss his proposal, for his claim is not addressed to the problem of word order freezing but to the verbal affix of Japanese. I assume that the flat structure is suitable for the purpose of my thesis.

Here, I used the OT mark (constraint name $\in o::*$) to model the preference of the grammatical functions. The OT-mark is an optimality theoretic constraint, which is ranked in the o-structure in XLE, which parses and generates LFG. The annotation $o::*$ states that the constraint is projected at o-structure (Crouch et al. 2008). Frank et al. (2001) suggest that the OT-mark is projected from the c-structure to the o-structure. Here, the OT-marks are stated under S at the c-structure. The linear order constraint 1 is called CANON-ORDER and states that the subject precedes the object. The competitive constraint 2 states that the object should be scrambled. The OT-ranking states that the canonical SOV order is favored over the scrambled one.

Consider the word order freezing:

- (21) Word Order Freezing in Japanese
 John=ga Yoko=ga suki.da
 John=NOM Yoko=NOM like.COP.PRES
 ‘John likes Yoko’

The subject is assigned to *John=ga*. The object is assigned to *Yoko=ga*, according to the linear order constraint CANON-ORDER. This results in f-structure (22)

- (22) F-structure for the sentence (21)

PRED	'like <SUBJ,OBJ >'				
SUBJ	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">PRED</td> <td style="padding-left: 5px;">'John'</td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">CASE</td> <td style="padding-left: 5px;">NOM</td> </tr> </table>	PRED	'John'	CASE	NOM
PRED	'John'				
CASE	NOM				
OBJ	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">PRED</td> <td style="padding-left: 5px;">'Yoko'</td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;">CASE</td> <td style="padding-left: 5px;">NOM</td> </tr> </table>	PRED	'Yoko'	CASE	NOM
PRED	'Yoko'				
CASE	NOM				

In the following, it is analyzed how the morphology contributes to the encoding of the grammatical function.

2.2. Morphology and Grammatical Function

This section examines the morphological means for the encoding of the grammatical function and considers the correlation between the case marking and the grammatical function. The surface case marking does not always correlate to a certain grammatical function. The subject is not always marked with the nominative case and the

(24) Subject test: morpheme *wa* (modified from Matsumoto (1996))

- a. Sensei=ga hon=o o-yomi-ni natta
Teacher=NOM book=ACC HON-read.PAST
'The teacher read the book'⁷
- b. Sensei=wa hon=o oyomi-ni natta
Teacher=TOP book=ACC HON-read.PAST
'The teacher read the book'

Furthermore, the dative case *ni* can mark the subject as well. Sentence (25) shows that the subject is marked with the nominative case, dative case or topic marker if the predicate is stative.⁸

(25) Subject test: the nominative object construction: a from Shibatani (1977)

- a. Sensei=ni eigo=ga o-deki-ni naru⁹
Teacher=DAT English=NOM HON-can.PRES
'The teacher can (handle) English'
- b. Sensei=ga eigo=ga o-deki-ni naru
Teacher=NOM English=NOM HON-can.PRES
'The teacher can (handle) English'
- c. Sensei=wa eigo=ga o-deki-ni naru
Teacher=TOP English=NOM HON-can.PRES
'The teacher can (handle) English'

This shows that the nominative case marks not only the subject but also the object if the predicate is stative. To diagnose the direct object, there are some object tests such as object agreement, case marking and relativization in LFG (Dalrymple 2001), but I assume that the insertion of *koto* (translated: *thing, fact*) is a good test for diagnosing the direct object in Japanese (e.g., Koizumi (2008), Kishimoto (2004))

⁷ The two sentences have the same meaning, but are different in the pragmatic structure.

⁸ Shibatani (2001) provides some finely categorized predicate group for the non-canonical case marking with the nominative object construction: (1) possession/existence (e.g., *aru/iru* 'have'), (2) psychological states (*sukida* 'fond'), (3) physiological states (*itai* 'painful'), (4) visual/audio perceptions (*mieru* 'visible'), (5) necessity/desiderative states (*hituyooda* 'necessary'), or (6) potentiality/ability (*hanaseru* 'can speak'). These verbs express a state rather than an activity. Here I use the term "stative predicate" for these verbs.

⁹ Lee (2003) discusses the honorific dative subject in Korean. She shows that it does not undergo scrambling. As for Japanese, I judge the sentence a (SOV) as low-acceptable, the OSV as unacceptable.

- (26) Object test: the insertion of *koto*: from Kishimoto (2004)
- a. John=ga Mary (-no-koto-)=o damashi.ta
 John=NOM Mary (-GEN-fact)=ACC deceive. PAST
 ‘John deceived Mary’
- b. *John-no-koto=ga Mary=o damashi.ta
 John-GEN-fact=NOM Mary=ACC deceive. PAST
 ‘John deceived Mary’
- c. Mary (-no-koto-)=o John=ga damashi.ta
 Mary (-GEN-fact)=ACC John=NOM deceive. PAST
 ‘John deceived Mary’

The sentences (26)a and (26)b show that the subject cannot have the phrase *no koto* be inserted but the object can. The phrase *no koto* is optional in the SOV and OSV sentences (26)a and (26)c. Using the phrase *no koto*, the nominative case marking can be tested as following:

- (27) Object test: the nominative object construction: a and b adopted from Kishimoto (2004)
- a. John=ga Mary (no-koto)=ga/*o suki.da
 John=NOM Mary (-GEN-fact-)=NOM/ACC like.COP.PRES
 ‘John likes Mary’
- b. John=ni Mary (no-koto)=ga/*o wakaranai
 John=DAT Mary (-GEN-fact-)=NOM/ACC understand.NEG.PRE
 ‘John does not understand Mary’

The nominative object is restricted to the stative-predicate, therefore the accusative case is excluded and annotated as /*o. The phrase *no koto* is optional for the object marking in the canonical word order (SOV) as shown by (27). It is however obligatory to appear in the non-canonical word order (OSV) as shown in the sentence (28).

- (28) Object test: the phrase *no koto* insertion: OSV ordering (Shibatani 2001)
- a. Mary no-koto=ga John=ga suki.da
 Mary (-GEN-fact-)=NOM John=NOM like.COP.PRES
 ‘John likes Mary’

- b. ? Mary=*ga* John=*ga* *suki.da*
 Mary=*NOM* John=*NOM* like.COP.PRES
 *‘John likes Mary’
 ‘Mary likes John’

Sentence (28) shows that the optionality of *koto*-insertion is restricted, dependent on the morpheme and word order.¹⁰

As for the morpheme *wa*, it can mark the direct object regardless of the predicate as illustrated in the sentence (29).

(29) Object test: the morpheme *wa*: a = not-stative, b = stative predicate

- a. John=*wa* Mary (-no-koto-)=*wa* *damashi.ta*
 John=*TOP* Mary (-GEN-fact)=*TOP*¹¹ deceive.PAST
 ‘John deceived Mary’
- b. John=*wa* Mary (-no-koto)=*wa* *suki.da*
 John=*TOP* Mary (-GEN-fact-)=*TOP* like.COP.PRES
 ‘John likes Mary’

To sum up, the Japanese main morphemes are illustrated in the table (30).

(30) Morphemes in Japanese

Clitic	Label	Grammatical Function
= <i>ga</i>	nominative	SUBJ, OBJ (stative-predicate)
= <i>o</i>	accusative	OBJ
= <i>ni</i> ¹²	dative	OBJ θ , SUBJ (stative-predicate)
= <i>wa</i>	topic	SUBJ, OBJ

I assume that the morpheme *wa* works as the indicator for the grammatical function SUBJ and OBJ in the language¹³ and there is no reason for it not to be defined as a case marker.¹⁴

¹⁰ The phrase *no koto* is indeed dependent on the animacy, too. Chapter 4 discusses this in detail in the OT-LFG account.

¹¹ Conventionally, the morpheme *wa* is glossed as TOP(ic). However, the *wa*-phrase is not always the topic. The *wa*-phrase here should be for example interpreted as focus. I will always gloss *wa* as TOP, regardless of its interpretation.

¹² The dative case *ni* marks the locative (OBL or ADJ) as well.

Supposing the morphemes *wa*, *ga* and *o* as case marker, we need some clarification on whether the case marking is an affix or clitic. The c-structure's terminal node should be a morphologically complete word following the lexical integrity principle in LFG.

- (31) Lexical Integrity (Bresnan, 2001: 92)
Morphologically complete words are leaves of the c-structure tree and each leaf corresponds to one and only one c-structure node.

The case marker as the affix means that it belongs to the morphological, lexical formation of the host (e.g., stem + affix), while the clitic has his own syntactic category and undergoes the syntactic rules at the c-structure. Vance (1993) provides some evidence for the Japanese case marker as a clitic, using the tests provided by Zwicky and Pullum (1983). They argue that it is true for the Japanese case marker to be “added to a host of almost any category” (Zwicky and Pullum 1983), which is typical for a clitic but not for an affix. He shows that the case marker can be hosted by the noun phrase (e.g., coordination sentence), verb (e.g., infinitive, gerund) and adverb as the sentence (32)a, (32)b, and (32)c respectively illustrate.

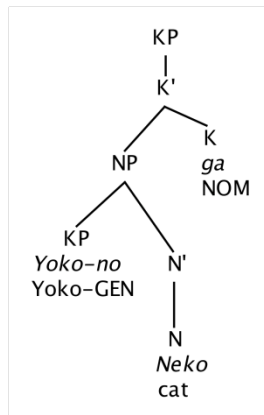
- (32) Clitic test: adopted from (Vance 1993)
- a. Taroo=*wa* [shinbun to hon]=*o* yondeiru.
Taroo=TOP newspaper and book=ACC read.GER.COP
'Taroo is reading a newspaper and a book.'
- b. Yamu=*o* enai koto datta.
Stop=ACC cannot thing be.PAST
'It couldn't be helped.'
- c. Kuwashiku=*wa* setsumei shinakatta.
in-detail=TOP explanation do.NEG.PAST
'(They) didn't explain in detail.'

¹³ Note that the morpheme *ga* is the nominative marker as a grammatical case marker, while the morpheme *wa* is the topic marker as “adverbial particle” in traditional Japanese linguistics. The morpheme *wa* is considered as an adverbial particle together with other particles such as *dake* (only), *made* (even) or *mo* (also), for its “adverbial” distribution: The morpheme *wa* can mark the noun phrase in addition to the marking of the semantic case and adverb, while the case markers do not. The topic marker *wa* is not so widely accepted to be the case marker as the morpheme *ga*. The morpheme *wa* is not discussed often as the case marker as shown in the summary of the case theory (Nishida 2004).

¹⁴ I do not follow the position that the “NP marking for definiteness, topic or focus have never been considered cases” (Haspelmath 2009). I will show in next chapter that the nominative case *ga* bears the feature [topic-] and the morpheme *wa* [topic +] in the notion of the information structure in Japanese.

In LFG, Butt and King (2000) provide also the coordination test for the clitic by which the Japanese case marker is diagnosed as the clitic in the sentence (32)a. I assume that the Japanese case marking is clitic and has the syntactic category K, following Butt and King (2005).

(33) Category K¹⁵



K is a functional category and the head of the maximal projection KP(KaseP). The category K contributes not only to the grammatical function information but the relevant feature around the case marking such as volitionality for the ergative case (Butt King 2004a, 2004b). The case can construct the feature by itself (“constructive case”) and is not only to be checked.¹⁶ The nominative object can appear only in the stative predicate. Such information can be constructed from the case marker. The topic marker *wa* can contribute to the information on the grammatical function (SUBJ, OBJ) and the information structure.¹⁷

¹⁵ The gloss is as follows:

Yoko=no neko=ga
 Yoko=GEN cat=NOM
 ‘Yoko’s cat (nominative)’

¹⁶ Traditionally, the case feature is checked by the predicate, and the predicate constructs the grammatical function and case feature (e.g., for example King 1995) as shown in the introduction. The description of the constructive case is developed in the work of Nordlinger (1998).

¹⁷ Sells (1998) provides the f-structural constraint on the case marker:

CAT ((↑GF), V)

That states that the larger f-structure of this grammatical function is restricted to the syntactic grammatical function V. Accordingly, the nominative marker is formulated as following:

CAT ((↑SUBJ), V)

CAT ((↑OBJ), V)

- (34) Lexical entry for K
ga: K
(↑CASE) = NOM
possibility 1: (SUBJ ↑)
possibility 2: (OBJ ↑)
 ((OBJ↑) V-TYP) = STATIVE
- wa: K
(↑CASE) = TOP
Possibility1: (SUBJ ↑)
 (↑i-str TOP)
- Possibility 2: (OBJ ↑)
- o: K
(↑CASE) = ACC
(OBJ ↑)

The case marker can provide information on the feature in i-structure in this formal description like (↑i-str TOP). The information structure will be discussed in detail in the next chapter.

3. Information Structure in Japanese

This chapter discusses the use of the morphemes *wa* and *ga* with respect to information structure. As the previous chapter discussed, both the topic marker *wa* and the nominative case *ga* correspond to the grammatical functions SUBJ and OBJ. The question now is what differentiates these two morphemes from each other. As Kuroda (2005) mentioned, there is some consensus among Japanese linguists (see Heycock (2007)) that the morpheme *wa* is the topic marker and *ga* the focus marker. Kuno (1973) claims in his influential work that *wa* bears the topic and contrastiveness and *ga* is used as the focus marker (“exhaustive reading” in his words) or is neutral. I agree with Kuno with regard to his distinction between *wa* and *ga*, if the morphemes are observed without interaction with the word order. As the realization of the information structure, there are some strategies such as morphological marking, word order or prosody. The strategy is different from language to language. I assume that the Japanese uses all three strategies to encode the pragmatic structure, not just morphology. This chapter concentrates on the morphology and linear word order but not the prosody. The reason is that it is crucial for the prosodic analysis to be based on the empirical data and an investigation of that will go beyond the scope of the master thesis.

Section 3.1 discusses the concept of information structure. Section 3.2 then provides an analysis of the morphological and syntactic use of the information structure of Japanese and introduces contrastive features. This is followed by its representation in LFG in Section 3.3.

3.1. Information Structure

Information Structure is the pragmatic structure produced during a discourse to deliver the message according to what the hearer knows and does not know. There is some terminological chaos in the term of information structure: Topic-comment, focus-presupposition, focus-background, theme-rheme, new/old, given/not-given, etc. I follow Valluduvu’s¹⁸ notion (1993) of the information structure (or his term “Information Packaging”) and assume that there are three major parts of the information structure: topic (“link”, using Valluduvu’s terms), focus, and background (“tail”, using his terms).

¹⁸ Valluduvu’s innovation lies in integrating the two relevant parts of the information structure, topic and focus, in one system.

(35) Information Structure (Valluduvi 1993)

S = {focus, presupposition}

Presupposition = {topic, background}

The sentence S can be divided into two major parts according to the informative value for the hearer. Focus constitutes the relevant new information for the hearer, while its complement, presupposition (in Valluduvi's term "ground") represents the old or known information. The presupposition is again separated into two parts: topic and background. The topic constitutes the prominent part of the presupposition and delivers the what-is-about of the sentence. The complement of topic, background is the residue of the topic and is less informative and prominent. According to Valluduvi (1995), the hearer's mind can be seen as the knowledge store or file card system, which is updated during the discourse. The speaker makes clear what the sentence is about (topic), so that the hearer locates the corresponding card in his knowledge store. The speaker indicates to the hearer what additional information the hearer should enter on this card (focus). The speaker thereby speculates as to what the he does or does not know or what they have already spoken about in their discourse or not. The mentioned, known entity can be realized as the topic or the background. The new, relevant information can be realized as the focus.

For example, sentence (36) shows how the information structure is partitioned in the Valluduvi's information component (1993):

(36) Information Structure from Valluduvi¹⁹(1993):

[T The boss] [F HATES] broccoli

GOTO (boss (x))(ENTER-SUBSTITUTE (V| X V broccoli))

Valluduvi does not explicitly give the context for the utterance, but the context could be uttered in the situation in which someone had said, "The boss loves broccoli". The boss is what-is-about in the sentence and realized as the sentence-initial position. The relevant information *hate* is realized as the focus by the prosodic mean in English. In this case, the hearer must revise and substitute what he already knows and enter the actual new information. As for the answer to the wh-question, the hearer just has to add the new

¹⁹ The annotation comes from the original work of Valluduvi. T is the abbreviation for the topic and F for the focus. The second line can be read as followings: Go to the card with the title "boss" and change the entry to "the boss HATES broccoli".

information to the card. The background is the third component of the information structure and corresponds here to *cauliflower*, which is already available to the interlocutors in the discourse. It is responsible for the decision whether new information is simply to be added to the record or old information must first be erased.

The next section further discusses the notion of information structure and its realization in Japanese.

3.2. Realization of Topic, Focus and Contrastiveness in Japanese

Before digging too deeply into the notion of topic, it is first necessary to discuss focus. Thereby, contrastiveness is discussed and combined with topic and focus.

3.2.1. Focus

The focus is the most salient information for the addressee in the speaker's estimation (Dik 1997). The focus is non-presupposed, relevant information in the sentence. A question-answer pair provides a useful test for the focus-ness. What is asked in the question and then answered is regarded as the relevant and new information in the current discourse.

(37) Focus test:

a. Question:

Dare=ga/*wa Kalifulawaa=ga sukina-no?
 Who=NOM/TOP Cauliflower=NOM like-Q
 'Who likes cauliflower?'

b. Answer possibility 1:

[Jyoushi=ga/wa]_F²⁰ Kalifulawaa=ga suki.da
 Boss=NOM/TOP Cauliflower=NOM like.COP.Pres
 'The BOSS likes cauliflower.'

c. Answer possibility 2:

Kalifulawaa=wa/#²¹ga [Jyoushi=ga/wa]_F suki.da
 Cauliflower=TOP/NOM Boss=NOM/TOP like.COP.Pres
 'The BOSS likes cauliflower.'

²⁰ I annotated the focused element with [_F].

²¹ The symbol # means that the sentence is grammatically correct but leads to a misinterpretation of the information structure.

There are two possible responses to the question “Who likes cauliflower?”. The meaning is almost the same but differs in the interpretation of the information structure. In both sentences, *Jyoushi* (translated as boss) is the focus; the questioned information is new and relevant. The questioner does not know who likes cauliflower in the previous sentence in the speaker’s estimation. The speaker and hearer do not share the information about who likes cauliflower. Observing the word order, the focused *Boss* can be preverbal ((37)c) or pre-preverbal ((37)b), regardless of the morphological marking. However, *cauliflower* is sensitive with regard to morphological marking and position. The entity cauliflower is already mentioned in the question and is therefore known and old. The known entity remains in the background in the sentence (37)b or topicalized in the sentence (37)c. It shows that the focus precedes the background and follows the topic. As for morphemes, *who* is marked with the nominative case marker *ga* and never with *wa*. The answer *Jyoushi* is marked with morpheme *ga* as well. This is the reason why the nominative case marker is often characterized as a focus marker. Nevertheless, the focused *Jyoushi* can be marked with the topic marker *wa* as well, as sentences (37)b and (37)c show. I follow Kuroda’s observation (Kuroda 2005) and claim that the focus can be marked with the topic marker *wa*. The alternation depends on whether the speaker considers other people who may love cauliflower. If the speaker thinks that the boss or some other common acquaintance such as friends or colleagues love cauliflower, the morpheme *wa* is used instead of the nominative marker *ga*. The nominative case *ga* implies that there is only one entity for the questioned gap in the speaker’s mind. This distinction is not new. The use of the morpheme *wa* has been discussed as comparativeness or contrastiveness and the nominative case marker as the exhaustive reading among the Japanese linguists for a long time (e.g., Kuno (1973)). The contrastiveness has been recently discussed not only with respect to Japanese, but also in other languages such as English (Erteschik-Shir 2007), Hungarian (Molnar 1998), and German (Molnar 1998, Buring 1999). Thereby, contrastiveness is regarded as compatible with the information structure and is combined with the topic and the focus. Regarding the property of contrastiveness, there is some consensus that it evokes or generates alternatives (e.g., see Valludivi and Vilkuna (1998) or Molnar (2001)).

Erteschik-Shir (2007) provides some remarks on the contrastiveness in the notion of the given/not-given and old/new.²² According to her account, the pair old/new is

²² There are some criticisms of the distinction between new and old, or given and not-given. The question pertains to the location of the boundary for the information structure. Is it simply the sentence? Reinhart (1972) demonstrate the following example: “Who did Felix praise? “He praised himself”. The answer should

decided with regard to previous mention of the referent in the discourse. The given and not-given pair is decided on whether the hearer has the referent in mind or not. The exhaustiveness (in her term “restrictive foci”) defines the set of the alternative entities just the same as the contrastiveness. Contrastiveness needs, however, “not be specified overtly” and can be derived from world knowledge, whereas exhaustiveness is subject to this constraint.

- (38) Contrastive focus (Erteschik-Shir 2007)
- a. Q: Which laundry did John wash, the white or the colored?
 - b. A: He washed the WHITE laundry
- (39) Exhaustive focus (or “restrictive focus”) from Erteschik-Shir(2007)
- a. Which one of his friends wants to meet John?
 - b. JANET wants to meet John.

White laundry is mentioned in the question and available in the discourse in the sentence (38)b, therefore the sentence is interpreted as “John washes the white laundry but not the colored laundry” in her account. Regarding the exhaustive reading, the entity Janet is not previously specified but given in the common ground as “one of his friends” in the sentence (39)b. The interpretation is different from contrastiveness in that it “does not mean that none of the their friends wants to meet him (John)”. Erteschik-Shir’s definition provides some distinction between contrastiveness and exhaustiveness, but it implies that if the entity is old then it signals contrastiveness, while if the entity is given but new, then it signals exhaustiveness. I assume that this is not the case for Japanese as shown in the sentences (37)b and (37)c. Rather, the entity is given for both contrastiveness and exhaustiveness but the old/new distinction is irrelevant for the contrastiveness. Givenness is more obscure to detect than oldness. The definition of the givenness is much clearer following the semantic approach of Rooth (1992). Accordingly, the focus entails the alternative set. Question (40) sets up this alternative.

- (40) Question: Which animal has a long nose?
- Focus set:
- {‘the elephant’, ‘the tapir’, ‘Proboscis monkey’ ..etc} has a long nose.

be new and identified as focus but “himself” is not new. As for givenness, it is difficult to say what is given in the mind of the interlocutors. Indeed, the givenness corresponds to the definiteness in Erteschik-Shir (2007). She gives the example such as “the moon” or “the composer” as the givenness.

The focus evokes the alternative focus entity to the presupposition “some animal which has a long nose”. The focused entity can be marked with the topic marker or the case marker in Japanese:

- (41) Question: Which animal has a long nose?
 Answer:
- a. A1: [Zou=WA]_F [hana=ga nagai]
 elephant=TOP nose=NOM long
 ‘The ELEPHANT has a long nose’
- b. A2: [Zou=GA]_F [hana=ga nagai]
 elephant=NOM nose=NOM long
 ‘The ELEPHANT has a long nose’

The sentence (41)a is interpreted as asserting that at least the elephant has a long nose. That hints at other animals, which may also have a long nose and shows “uncertainty” or “incompleteness” in the speaker’s mind (Tomioka 2009). The speaker is unsure of which other animals have a long nose and does not exclude the possibility of alternatives to the elephant. A complete, certain answer would entail only one entity (the elephant) in the speaker’s mind, resulting in exhaustivity of the focus with the nominative case such as in answer (41)b. Molnar (2001) defines the exclusional property of these two focus types as “weak exclusion” (“some-exclusion”) and “strong exclusion” (“all-exclusion”).²³ The weak exclusion means that some alternatives are excluded from the set but not all, corresponding to the interpretation for the contrastive focus with *wa*-marking. The strong exclusion means that all alternatives are excluded from the set, corresponding to the exhaustive focus reading with the *ga*-marking.

- (42) Two types of Exclusion, two types of focus
- a. Some-exclusion = [+contrastiveness] => Contrastive Focus *wa*
- b. All-exclusion = [-contrastiveness] => (Exhaustive) Focus *ga*

²³Originally, Molnar discusses the contrastive topic, not the contrastive focus. She defines the focus as exclusive and exhaustive, yet defines the contrastive topic as exclusive and not exhaustive. However, her description corresponds to the contrastive focus as well.

I define here the information structure feature, adopting the feature space from Valluduvi and Vilkuina (1998).²⁴

(43) Features: Information Structure

	Contrastive [+C]	Not-contrastive [-C]
Topic [+T]	Aboutness topic	Contrastive Topic
Focus [-T]	Focus <i>¬WA & precedence of BCK</i>	Contrastive Focus <i>WA & precedence of BCK</i>

The feature is divided into topic [+T] (topic) and focus [-T] and combined on the basis of whether the contrastiveness is present [+C] or not [-C]. I follow Valluduvi and Vilkuina (1998) and assume that the focus [-T] constitutes the part of the sentence which is possible to update about the topic. The focus is realized by the morphological (not =*wa* as *¬wa*) and syntactic (the focus position: between the topic and the background) and the contrastive focus is articulated by using the morpheme WA (with accent) and by preceding the background. The realization of the topic is examined in next subsection.

Before discussing the topic and contrastive topic, it should be noted that a four-way distinction of the information structure has been proposed in LFG. Choi (1996) proposed following distinction based on the Valluduvi’s information component system:

(44) Choi’s Information Structure Feature (Choi 1996)

	+PROM	-PROM
-NEW	TOPIC	BACKGROUND
+NEW	CONTRASTIVE FOCUS	FOCUS

Choi divides the presupposition (-NEW) into the prominent part topic and unprominent part background. The focus (+NEW) is divided into the prominent component contrastive focus and the unprominent component focus. Please also note that there is no contrastive topic for her system. She argues that the topic and contrastive focus are both prominent because both are fronted in English:

²⁴ Originally, Valluduvi and Vilkuina define the information structure with [\pm Rh(eme)]. However, I use theme just for the sake of the topic/non-topic distinction. K is used for the contrastiveness in the work of Valluduvi & Valkuna (1998). BCK stands for background.

(45) Fronting in English (Choi 1996)

- a. Topic
BEANS, I don't LIKE
- b. Contrastive focus
FIDO they named their dog
- c. Focus
They named their dog FIDO

The topic and contrastive differ from each other in that the contrastive focus compares strongly the alternatives in the set but the topic does not. It is unclear what determines the prominence. The system is oriented to the linguistic realization but not the property of the information structure.

Butt and King (1996) provide other distinctions. The features are the same as in Choi's system. Their categorization differs from Choi's in that new information is separated into the completive information and the focus.

(46) Butt & King's Information Structure Feature (Butt and King 1996)

	+PROM	-PROM
-NEW	TOPIC	BACKGROUND
+NEW	FOCUS	COMPLETIVE INFORMATION

The *completive information* used here has not been discussed yet. It is new entity in the discourse and is less prominent or informative than the focus. The Urdu data shows why the new information is divided into these two parts in their system.

(47) Completive Information in Urdu from Butt and King (1996)

- a. Question: Where is Nadya coming from?
- b. Answer 1:
Naadyaa to abhii tofii [bazaar=me] F
Nadya.F.Nom indeed just now toffee.F.Nom market.M=in
xarid rah-ii t-ii
buy Stat.F.Sg was-F.Sg
'Nadya was just buying toffee AT THE MARKET'
- c. Answer 2: # Naadyaa to abhii [bazaar=me] F xarid rah-ii t-ii [tofii]_{back}
'Nadya was just buying toffee AT THE MARKET'

The sentence (47)b shows that the relevant new information *bazaar=me* is positioned at the preverbal phase. The new but uninformative entity *toffee* is positioned between the topic and focus. The *toffee* is however not allowed to be at the postverbal phrase, which is the position of the background in Urdu. The sentence (47)c shows that the *toffee* cannot be positioned postverbal like the background but not as prominent or informative as the asked information *bazaar=me*. The position of the completive information is at *toffee*, which precedes the focus.

Japanese shows some similarity to the Urdu data. The completive information precedes the real focus as shown in the sentence (48)b.

- (48) The completive information in Japanese
- a. Question: Where is Nadya coming from?
 - b. Answer 1:

Nadia=*wa* choudo ima tofee=*o* ichiba=*de*
 Nadya=*TOP* just now toffee=*ACC* market=*LOC*
 katte.iru.tokoro.da
 Buy.ASP.ASP.COP
 ‘Nadya was just buying toffee AT THE MARKET’
 - c. Answer 2:

#Nadia=*wa* choudo ima ichiba=*de* tofee=*o*
 Nadya=*TOP* just now market=*LOC* toffee=*ACC*
 katte.iru.tokoro.da
 Buy.ASP.ASP.COP
 ‘Nadya was just buying toffee AT THE MARKET’

The sentence (48) shows that the informative new element *ichiba=de* is realized as the focus that is preceded by the topic and the completive information *toffee=o*. The preverbal phrase is indeed the position for the background in Japanese as discussed before, therefore the sentence (48)c is odd. Clearly, the *toffee* corresponds to the completive information in Urdu. I revise the linear order of the discourse function in Japanese as followings:

(49) Linear order of the discourse function: case marking

Completive (Discourse neutral)	<	Focus	<	Background (Discourse neutral)
+NEW		+NEW		-NEW
-PROM		+PROM		-PROM

The informative, prominent component is positioned in the middle of the components and the new information comes first. The completive and backgrounded information are not prominent and do not correspond really to the focus.

The topic and its realization are discussed in the following section. The contrastiveness is thereby more precisely examined.

3.2.2. Topic

The topic is the aboutness of the sentence. More precisely, an entity is the topic, if the speaker “intends to increase the addressee’s knowledge about, request information about”(Guntel 1988). Pragmatically, the topic is known, old, and relevant information for the hearer, as Valluduvi (1993) discussed. Semantically, the topic can be tested based on answering questions like “tell me about the boss,” as Reinhart (1981) proposed. Syntactically, the topic can be tested with “as for” or “speaking of” phrases (Erteschik-Shir 2007). As for Japanese, it is well known that the morpheme *wa* marks the aboutness of the sentence.

- (50) Topic-comment structure
- | | | |
|-----------------------|----------|---------------------------|
| [Zou=wa] _T | [hana=ga | nagai] _{Comment} |
| Elephant=TOP | nose=NOM | long |
- ‘As for elephant, the nose is long’

The sentence (50) is a typical sentence in the topic prominent language (Li and Thompson 1976). The sentence is clearly about the elephant, whose nose is long. The aboutness-condition on the topic corresponds to the first property (51) of the topic in Strawson’s definition of the topic:

- (51) Strawson's property of the topic²⁵ cit. from Erteschik-Shir (2007)
- The topic is what a statement is about
 - The topic is used to invoke knowledge in the possession of an audience
 - The statement is assessed as putative information about its topic

The noun phrase *elephant* bears the topic feature and is articulated by fronting and marking with the topic marker. The noun phrase *nose* is marked with the nominative case marker and it constructs the comment together with the predicate adjective.

The question is now whether the topic is always marked with the morpheme *wa* and at the clause-initial position in Japanese.

- (52) Topic position the same sentence a, b, c from the focus test (37)²⁶
- Q: Dare=ga Kalifulawaa=ga sukina-no?
Who=NOM Cauliflower=NOM like-Q
'Who likes cauliflower?'
 - A1: [Jyoushi=ga/wa]_F [Kalifulawaa=ga]_{BCK} suki.da
Boss=NOM/TOP Cauliflower=NOM like.COP.Pres
'The BOSS likes cauliflower.'
 - A2: [Kalifulawaa=wa]_T [Jyoushi=ga/wa]_F suki.da
Cauliflower=TOP Boss=NOM/TOP like.COP.Pres
'The BOSS likes cauliflower.'
 - A3: #Jyoushi=ga/wa [Kalifulawaa=wa]_F suki.da
Boss=NOM/TOP Cauliflower=TOP like.COP.Pres
* 'The BOSS likes cauliflower.'
'The boss likes CAULIFLOWER.'
 - A4: #[Kalifulawaa=ga Jyoushi=ga/wa]_F suki.da
Cauliflower=NOM Boss=NOM/TOP like.COP.Pres
* 'The BOSS likes cauliflower.'
'The BOSS likes CAULIFLOWER.'

²⁵ The second property corresponds to the state of the knowledge store in the hearer's mind in Valludivi's system. The third property of the topic is concerned with the truth-value of a sentence. The topic is the pivot where the truth-value of the sentence assigns.

²⁶ I annotate the topic as [_T] and the background as [_{BCK}].

There are four alternative responses to the question “who likes cauliflower”. The *cauliflower* is topicalized and is marked with the topic marker in sentence (52)c. As expected, the *cauliflower* is the what-is-about in the sentence and bears the topic property. In contrast, the sentence (52)d is odd as the answer to the question (52)a. The answer would be appropriate for the question such as “what does the boss like?” or “does the boss like the broccoli?” The sentence is interpreted as meaning that the boss at least likes cauliflower. *Cauliflower* bears the contrastiveness (some-exclusion) and the focus property (the informative part of the sentence) at the preverbal position. The sentence (52)e shows that *cauliflower* is syntactically topicalized but marked with the case marker. It is an odd sentence, and could likely be uttered as the response to a question such as “what are you talking about?” The sentence is a kind of all focus sentence. This results in articulating the topic by marking with *wa* and fronting syntactically. Neeleman et al. (2009) come to the same result that the topic is realized by the mean of the fronting and the *wa*-marking.

However, there is some difficulty in assigning the topic if an adverbial phrase is inserted into a transitive sentence. Shibatani (1990) claims that adverbial elements are fronted and marked with the topic marker but are not real topics.

- (53) Adverbial-*wa* phrase cit. Shibatani (1990)
- a. *Kyou=wa* [tenki=*ga* ii].
 Today=TOP weather=NOM good
 ‘Today is such that the weather is good.’
- b. *Kyou=wa* [boku=*ga* [e] ryouri-siyou].
 today=TOP I=NOM cook-do.VOL
 ‘Today, I will cook.’

Kyou=WA will be interpreted as contrastive (“today, but not tomorrow”), if the accent falls on the *wa*-phrase. Fronted *Kyou=wa* receives some emphasis and is interpreted as ‘I will cook today, what is special’, if it is not accented. Interestingly, Shibatani (1990) claims thereby that unaccented *today* is the topic in the sentence (53)a, but not in (53)b. The sentence (53)a describes the property of today and the sentence (53)b does not. Shibatani defines the adverbial phrase *kyou=wa* ((53)b) as the “stylistic” topic and describes “a simple and slight emphasis”. Clearly, *today* is not as topic-like as the real topic and does not fulfill the aboutness condition, but is more prominent than the

background. I will categorize the phrase *kyou=wa* ((53)b) as discourse neutral,²⁷ which acts neither as the topic nor the focus and does not fulfill the aboutness-condition.²⁸

Neeleman et.al (2009) also discuss the *wa*-phrase, which acts neither as the topic nor contrastive focus:

- (54) *wa*-phrase neither in the topic nor in the focus position cit. from Neeleman et al. (2009)
- a. Q: Who did the dog bite?
- b. A1: [sono inu=wa]_T [John=o]_F kinoo kande.simatta
that dog=TOP John=ACC yesterday bite.ASP.PAST
‘The dog bit JOHN yesterday.’
- c. A2: ²⁹[John=o]_F sono inu=wa kinoo kande.simatta.
John=ACC that dog=TOP yesterday bite.ASP.PAST
‘The dog bit JOHN yesterday.’

Inu=wa ‘dog=TOP’ is positioned at the sentence initial position ((54)b) and in the middle of the sentence ((54)c). The first sentence corresponds perfectly to the topic sentence and would be the preferred response to the aboutness-condition sentence “tell me about the dog”. In the later sentence, the *dog* is marked with the topic marker but does not really correspond to the topic. The dog is hearer-old but not as prominent as the topic. The *dog* is “discourse anaphoric” in their term. They do not define the phrase *inu=wa* either as the topic or the background. Obviously, *inu=wa* is not as topic-like as the real topic (fronted and *wa*-marking) but too prominent to be categorized as the background. I will categorize it discourse neutral, which is associated neither with the topic nor with the focus. The discourse neutral component is governed by the elsewhere principle and is assigned regardless of the old/new or given/not-given distinction. The discourse neutral

²⁷ Valluduvi’s background is too narrow: The background is hearer-old in his definition.

²⁸ The aboutness condition can be examined for the non-*wa* phrase. Fry and Kaufmann (1998) claim that the *ga*-phrase can act sometimes as the topic.

- a. John=wa kaetta
John=TOP return.PAST
‘John went home’
- b. John=ga kaetta
John=NOM return.PAST
‘John went home’

They don’t give the context explicitly but the sentence b is not an acceptable response to the question: “tell me about John”. The aboutness-condition is not fulfilled. I assume that the topic marking is mandatory for the topic assignment.

²⁹ The answer b is a grammatical correct sentence, but the question should be “who bites the dog yesterday?”.

component includes the background and the completive information. Sentence (54)d shows that this discourse neutral component is not positioned at the preverbal phrase.³⁰

(56)

- d. #John=o kinoo sono [inu=wa]_F kande.simatta.
 John=ACC yesterday that dog=TOP bite.ASP.PAST
 ‘The DOG bit John yesterday.’

Inu=wa receives the contrastive focus interpretation in the sentence (54)d. That indicates that the adverbial element *yesterday* blocks the contrastive interpretation of *inu=wa* in the sentence (54)c. The preverbal position is reserved for the contrastiveness, if the phrase is marked with *wa*. Let’s return to the cauliflower sentence. In the sentence (52)b, focused *Jyoushi=wa* ‘boss=TOP’ is not positioned preverbally but clause-initially. It still receives the contrastive focus interpretation. That indicates that the contrastive focus is not always realized by the preverbal position and the *wa*-marking. Rather, the contrastiveness is assigned to the initial or preverbal phrase and the linear order assigns the focus feature to the *wa*-phrase.

(55) Contrastive position in Japanese: topic marking

±Contrastive	< ¬ Contrastive	< + Contrastive	Verb
Some-exclusion	All-exclusion	Some-exclusion	

The sentence initial contrastiveness is dependent on the accent as the elephant sentence shows. With regard to the preverbal *wa*-phrase, it is prosodically conditional that

³⁰ Kaiser (1998) suggests that the post-verbal phrase acts as the background in Japanese. Written Japanese is a strict verb-final language. In the spoken language, however, some phrases can follow the verb. Almost every category can be extraposed behind the verb, with the exception of the head element (see Sells 1998). The postposed phrase is not always hearer-old but is always non-relevant information, according to Kaiser. It seems to me that the postposition is possible for all three information components, focus (a), topic (b) and discourse neutral(c).

a. Q: ‘How is John going?’

A: John=wa [shibaraku tabishiteita rashii yo, Amerika=ni]_F.
 John=TOP for-a-while travel.ASP it-is-said EMP America=LOC
 ‘John was traveling in America for a while’

b. Q: what does John like?

A: sushi=ga suki.da yo ne, [John=wa]_{TOP}
 Sushi=NOM like.COP EMP EMP John=TOP
 ‘John likes sushi’

c. Q: ‘What did John buy yesterday?’

A: John=wa sakana=o katta rashii yo, [kinou=wa]_{DN}
 John=TOP fish=ACC buy.PAST it-is-said EMP yesterday=TOP
 ‘John bought fish yesterday’

the accent falls on the preverbal phrase,³¹ which receives the contrastive reading. The dog sentence (54)c shows that *inu=wa* ‘dog=TOP’ remains not at the preverbal phrase and *yesterday* blocks *inu=wa* to be assigned to the contrastiveness at the preverbal phrase.

What about the contrastive topic? What is the difference from the topic? Crosslinguistically, the contrastive topic ascribes to the semantic or pragmatic aboutness (e.g., Erteschik-Shir (2007), Valluduvu (1993)) or the intonational contour (e.g., B-accent for English, e.g., Büring (2007)) or the topicalization or sentence-initial position (e.g., Molnar (2001), Umbach (2001), Büring (1999)) as follows:

- (56) Contrastive topic with the aboutness-condition from Erteschik-Shir(2007)
- a. Tell me about brothers John and Bill.
 - b. JOHN_{CT}³² is the smart one.
- (57) Contrastive Topic with B-accent in English from(Büring 2007)
- a. What did the pop stars wear?
 - b. The FEmale_{CT} pop stars wore CAftansF.
- (58) Contrastive Topic with the syntactic property from Umbach (2001)
- a. Adam: What did the children do today?
 - b. Ben: The small_{CT} children stayed at HOME and/but the bigger_{CT} ones went to the ZOO.

All three answers imply that there is an alternation in the speaker’s mind, explicitly or not: from John to Bill, from the female pop star to the male pop star, from the small children to the bigger children. These entities are the aboutness in the sentence, too.

How does Japanese handle contrastive topic? The same sentence as sentence (59) can be interpreted as the contrastive topic or aboutness topic without the context. The question is different in both interpretations:

³¹ According to Ishihara (2000a), the main stress (i.e., nuclear stress) falls on the preverbal phrase in Japanese. Naturally, the question arises what happens to the focus if it is not preverbally positioned. He argued that the additional accent creates a higher prosodic peak than the main stress and can be assigned to any element in a sentence, whereas the main stress falls always on the preverbal phrase. The element after the additionally accented phrase is deaccented and may correspond to the discourse neutral component (i.e., background). I assume that the contrastive focus is always assigned to the preverbal phrase as long as the phrase is marked with *wa*, but possible to be elsewhere in the sentence. I did not discuss the prosodic phrasing in this thesis but it is necessary to investigate in it in detail to proceed with the discussion.

³² CT is the abbreviation for the contrastive topic.

- (59) Aboutness topic and contrastive topic
- a. A1: [Zou=wa]_T [hana=ga nagai]_{Comment}
 Elephant=TOP nose=NOM long
 ‘The elephant has a long nose’
 Q2: “Which body part of the elephant is long?”
- b. A2: [Zou=WA³³]_T [hana=ga nagai]_{Comment}
 Elephant=TOP nose=NOM long
 ‘The elephant has a long nose’
 Q2: “What body part of the monkey is long?”

Sentence (59)a is uttered in response to the question “Which body part of the elephant is long?” The fronted *elephant=wa* bears the topic feature because of the aboutness and the entity previously mentioned. The nose is the element in question and is focused in the sentence. The sentence (59)b is the answer in the context: “what body part of the monkey is long?”. This answer is about the elephant, which is new in a way and should be the focus, but the information is irrelevant to the questioner. The answer does not really respond to the question. We can either define the elephant as the contrastive focus for the sake of the new entity (but not irrelevant information) or the contrastive topic for the sake of the aboutness condition (but the entity is not old). In the speaker’s mind, the elephant is given for the both cases. The sentence (59)b is interpreted as meaning that at least the elephant has a long nose. That implies that the speaker does not know exactly what body part of the monkey is long or does know but changes the topic to the elephant, which has a long nose.

To account for contrastive topic, I follow Büring (1999) and assume that the topic evokes alternative sets as well. Presupposed that the elephant is the contrastive topic, sentence (59)b generates the topic set:

³³ The capitalization annotation indicates the pitch accent. I follow and assume that contrastive topic receives the focal accent (Tomioka 2009), while the aboutness-topic does not (Komagata, 2001). Indeed, Tomioka proposed two types of contrastive topic: ZOU=wa and zou=WA.

(60) Topic set

- {The elephant has ‘a long nose’, the elephant has ‘a long neck’...}
- {The monkey has ‘a long nose’, the monkey has ‘a long neck’...}
- {The mouse has ‘a long nose’, the mouse has ‘a long neck’...}
- {The stork has ‘a long nose’, the stork has ‘a long neck’...}

The topic set is the set of the focus sets, in which the topic can be replaced in the question by other entities such as the monkey and the mouse. The focus set for the original answer is one from the topic set: {The monkey ‘has a long nose’, the monkey ‘has a long neck’...}. Büring’s approach provides an account of why sentence (59)b evokes some alternative entities, signaling another question to the hearer such as “What about the giraffe?” (“Residue Topic” in Büring (2007)). It bears the same property as the contrastive focus in that the alternative is generated.

To sum up: the realization of the information structure with the topic marker is illustrated as following:

(61) Linear order of discourse functions with the topic marking³⁴

Topic	<	Discourse Neutral	<	Contrastive Focus	Verb
+ PROM		-PROM		+PROM	
+ABOUT		-ABOUT		-ABOUT	
±CONTRA				+CONTRA	

(62) Linear order of the discourse function with morphemes

Topic	<	Completive (Discourse Neutral)	<	Focus	<	Background (Discourse Neutral)	Verb
± NEW		+NEW		+NEW		-NEW	
+PROM		-PROM		+PROM		-PROM	
+ABOUT		-ABOUT		-ABOUT		-ABOUT	
		<i>Case marking</i>		-CONTRA			
±CONTRA		<i>Topic marking</i>		+CONTRA			
				(Preverbal)			

³⁴ ABOUT is the abbreviation for aboutness, CONTRA of contrastiveness.

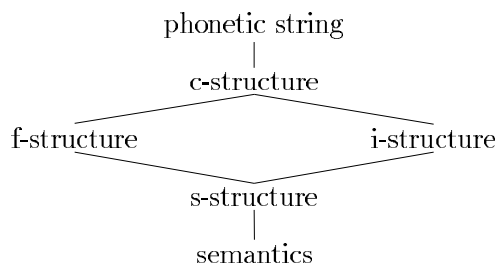
This table shows that case markers are used for the focus and discourse neutral component. The topic marker is used for all four components and for the contrastiveness (sentence initial and preverbal).

The analysis now converts in the formal description of LFG.

3.3. Information Structure in LFG

In LFG, the topic and focus are traditionally encoded in the f-structure (Bresnan 2001). Discourse functions such as TOP(ic) and FOC(us) are integrated at the f-structure and bear the appropriate grammatical function in the PRED(icate)³⁵ (Bresnan 2001). The discourse function is often used in addition to an unspecified grammatical function, in order to model the long distance dependency (e.g., the topicalization, wh-question, or relative clause). Discourse function was originally developed to capture topic prominent languages. However, some serious problems arise if the information structure is represented at the f-structure. The first problem arises when assigning a predicate to a discourse function, as King (1997) pointed out. As we have already seen in the sentence, “The boss HATES broccoli”, not only do the arguments bear the discourse function, but so does the predicate. The problem is that the same discourse function is assigned to the arguments, if its predicate is assigned to a discourse function (King 1997). Furthermore, the discourse function does not always correspond perfectly to the syntactic constituent. Confronting such problems, King (1997) proposed the separate representation of I(nformation)-structure, which is projected from the c-structure and can be mapped onto the s(ematic)-structure as illustrated:

(63) I-structure from King (1997)

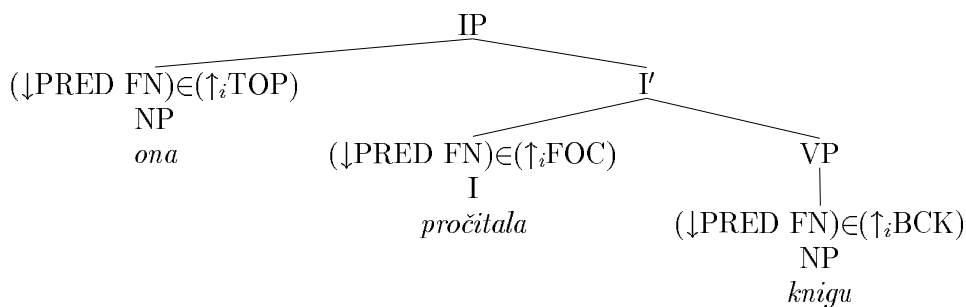


³⁵ According to extended coherence (Bresnan 2001), not only the argument function but also the discourse function must be bound.

King (1997) proposed discourse configurationality to be represented at the c-structure. The Russian sentence (64) is represented at the c-structure, f-structure, and i-structure.

- (64) Russian (King 1997).
 Ona PROCITALA knigu
 She read the book
 ‘She READ the book’
 focus = contrastive on ‘read’

- (65) C-structure, f-structure, and i-structure in Russian (King 1997).



F-structure

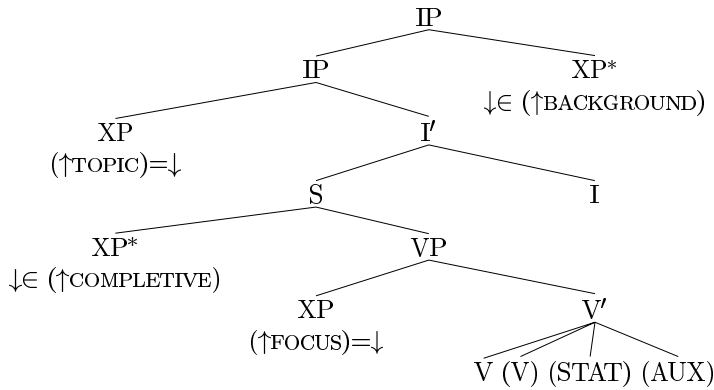
PRED	‘read<SUBJ,OBJ>’
SUBJ	[PRED ‘she’]
OBJ	[PRED ‘book’]

I-structure

TOP	{SHE}
FOC	{READ}
BCK	{BOOK}

The fixed phrase position assigns the discourse function at the c-structure. The discourse function is projected from the c-structure to the i-structure as the symbol $\uparrow i$ indicates at the c-structure. The parenthetical \downarrow PRED FN takes out the form of the predicate and separate it from its arguments so that at i-structure, one can see which words are in TOP, FOC, etc. The symbol \in indicates that more than one value can be inserted to the value. That corresponds to the representation at the i-structure with the symbol $\{\}$. All-focus sentences have, for example, more than one value for the focus attribute. The i-structure has three attributes: TOPic, FOCus, and BCK(background). In King’s approach information structure can be encoded in the annotation of the c-structure by default. Butt and King(1996) propose a c-structure for discourse configurational languages such Urdu and Turkish as follows:

(66) Discourse Configurability in LFG from Butt & King(1996)



The specifier position is the location for discourse functions such as topic, focus, and background in the classical LFG framework. The two languages are not strictly verb-final, and both allow the background to be positioned postverbally, with the preverbal position being reserved for the focus. Japanese is, however, a verb-final language and the preverbal position is not reserved exclusively for the focus. As already discussed, the discourse neutral component can be positioned between the focus and the verb. The lack of the preverbal focus position suggests that the assignment of discourse is carried out by the linear order.³⁶ Therefore the c-structure remains flat without the VP-node without its specific discourse function. There is neither a specific discourse nor grammatical function under the S-node.

³⁶ Butt and King (1998) proposed the p(honological)-structure to represent the prosodic phrasing (e.g. intonational phrase, phonological phrase and prosodic word). They argued that the p-structure is projected from the c-structure which entails the syntactic constituency. I follow their approach and assume that some extra phrasing is needed to constraint the contrastiveness at the preverbal phrase in Japanese:

<p>C-structure: S → XP*</p> <p style="padding-left: 2em;">p::* \$ (p:: M* ProsW)</p>	<p>YP</p> <p style="padding-left: 2em;">p::* \$ (p:: M* ProsW)</p> <p style="padding-left: 2em;">↓ = (↑i-str CONTRA)</p> <p style="padding-left: 2em;">(↓ CASE) = TOP</p>	<p>V</p> <p style="padding-left: 2em;">p::* \$ (p:: M* ProsW)</p>
<p>P-structure: S → XP*</p> <p style="padding-left: 2em;">p::* \$ p:: M*</p> <p>ZP → YP</p> <p style="padding-left: 2em;">p::* \$ p:: M*</p>	<p>ZP</p> <p style="padding-left: 2em;">p::* \$ (p:: M* ProsP)</p> <p>V</p> <p style="padding-left: 2em;">p::* \$ p:: M*</p>	

The annotation p::* \$ (p:: M* ProsW) states for example that this node corresponds to the mother's node, which has the attribute Pros(odic)W(ord). ProsP stands for prosodic phrase. YP is associated with the contrastiveness at the i-structure and restricted to the topic marker. P-structure is here built in bottom-up and the prosodic phrase is added, in which the contrastiveness is assigned. Probably, it is necessary to have more prosodic layers for Japanese (see Uechi 1998).

(67) Discourse Function at the C-structure in Japanese

S → KP* V
 (↑GF) = ↓ ↑ = ↓
 (↑DF) = ↓

GF order preference: SUBJ < OBJ

Pragmatic order constraint:

S: TOP < FOC

DF is the abbreviation for a discourse function (e.g., TOPic, FOCus, BaCKground). Precedence in DF is represented by the symbol <, which means: precede. I follow King's approach to model the information structure at the i-structure in Japanese and use the attribute TOP, FOC, DN (discourse neutral component) and CONTRA(stiveness). We need some constraints to ensure that the discourse function is assigned according to the linear order. The pragmatic order constraint states that the topic argument precedes the focus argument.

Besides the linear order, Japanese has some morphological means of encoding the discourse function. As already demonstrated, the morpheme *wa* can bear the aboutness topic, contrastive topic, and contrastive focus features:

(68) Lexical entry for the topic marker *wa*

wa: K

(↑CASE) = TOP

(↑SUBJ)|(↑OBJ)

possibility 1: (↑i-str TOP)

possibility 2: (↑i-str DN)

The first line states that the morpheme *wa* can bear the subject or the direct object in the f-structure. Possibility 1 shows that the discourse function is projected into the i-structure and bears the attributes TOP(ic). Possibility 2 states that the morpheme *wa* is discourse neutral. As for the topic and DN, we must regulate the presence of the focus, following Valluduvu's insight that all sentences have a focus.

((TOP ↑i-str) FOC)³⁷

((DN ↑i-str) FOC)

The morpheme *wa* can be accented and needs another lexical entry as *WA*:

(69) The lexical entry for the topic marker WA

WA: K

(↑CASE) = TOP

(↑SUBJ)|(↑OBJ)

(↑i-str CONTRA)

possibility 1: (↑i-str TOP)

possibility 2: (↑i-str FOC)

The lexical entry states that the topic marker assigns the contrastive topic (possibility1) and the contrastive focus (possibility 2).

The case marker can mark the focus. The attribute CONTRA remains negative.

(70) Focus feature for the nominative and accusative cases

GA: K

(↑CASE) = NOM

(↑SUBJ)|(↑OBJ)

(↑i-str FOC)

ga: K

(↑CASE) = NOM

(↑SUBJ)|(↑OBJ)

(↑i-str DN)

O: K

(↑CASE) = ACC

(↑OBJ)

(↑i-str FOC)

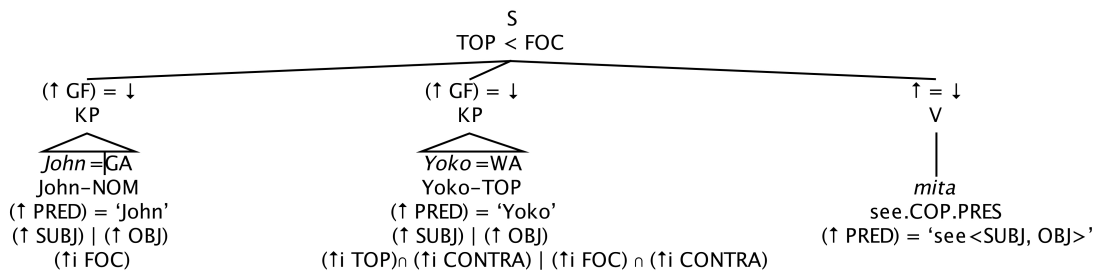
³⁷ I follow the annotation of the constructive case (Nordlinger, 1998).

- o: K
- (↑CASE) = ACC
- (↑OBJ)
- (↑i-str DN)

Consider the sentence in Japanese.

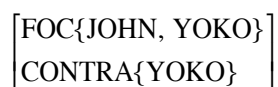
- (71) 'John saw Yoko'
 John=GA Yoko=WA suki.da
 John=NOM Yoko=TOP see.PAST
 'JOHN saw YOKO'

- (72) C-structure for the sentence (71)



The focus is assigned to the *John=GA*. The lexicon provides more than one possible discourse function for Yoko. Morphologically, Yoko can bear the contrastive topic or contrastive focus; the pragmatic order constraint states that the topic precedes the focus. It discards the topic feature. The contrastive focus is assigned to the preverbal *Yoko=WA*. This results in the i-structure (73).

- (73) I-structure for the sentence (71)



To sum up: the topic is assigned at the clause-initial position and marked with *wa*. It must be filled with the aboutness condition. The topic marker *wa* can also mark the contrastive focus, which does not exclude all the alternatives (“some-exclusion”) and differs from the (exhaustive) focus (“all-exclusion”). The all exclusive focus is marked with case markers.

The third component is discourse neutral, which associates neither with the topic nor the focus. It is realized with case markers. Topic, focus and discourse neutral component are assigned to the relative linear order, which differs from the preverbal contrastive focus assignment. There is no evidence for the focus position in Japanese, since the discourse neutral component follows the focus and can be positioned preverbally. Accordingly, I proposed a discourse flat structure in Japanese.

4. OT-LFG

This chapter discusses the interaction between morphemes, word order, semantics (e.g., animacy), and information structure in the OT-LFG approach. As discussed in previous chapters, word order in Japanese is relatively free, allowing SOV- and OSV-ordering in transitive sentences. This freedom of word order is, however, limited to SOV-ordering if the morphology does not work as a linker to the grammatical function. On the other hand, word order is optimized to articulate information structure. To account for such restriction and optimization of word order, I assume that it is crucial to observe both functions together, which the word order performs in the language. Previous chapters discuss and describe morphology, word order, and information structure in their own linguistic representations in the classical LFG. Now the findings should be combined to come together on the same stage, to observe how the various constraints work together and affect each other. To model the interaction, I assume that classical LFG must be supplemented by OT as a sub-framework so that multiple constraints compete with each other on a single stage. Fortunately, LFG has some tradition in that direction. Classical LFG is combined with OT (Prince and Smolensky 2004) and conceptualized as OT-LFG (Bresnan 1996). OT-LFG is indeed a happy mixture of the descriptive, formal LFG, and procedural, conceptualized, explicative OT. The biggest difference between LFG and OT-LFG is the property of the grammatical principle. The classical LFG describes the grammar as principles which should not be violated. In OT-LFG, the constraints are essentially violable and therefore weightable. Grammaticality is ensured by the ranking order of relevant constraints for a particular language.

This chapter is organized in two parts: the presumption and analysis of the interaction between the word order and morphology. The first part is considered with the basic assumption that the OT-LFG sets up the input (section 1) with the specification of the f-structure by linking from the a(argument)-structure (section 2). The second part is dedicated to the competition between morphology and word order among the constraints on the identification of the grammatical function (section 3) and those on the realization of the information structure (section 4).

4.1. Input

The production-based OT-LFG is grounded in the procedural conception that the grammar GEN(erator) is considered to be a set of inviolable universal principles which generates a set of candidate forms for a certain meaning. The candidates are considered language-indifferent and possible for some language.³⁸ The candidate set is evaluated by the violable, weighable constraints, all of which are considered universal, with variations in their ranking from language to language. The winning candidate set in the competition is the one that violates the fewest highly ranked constraints and is considered as output. The relevant semantic information is thereby given as input at the starting point.

In the OT-LFG, the input contains an underspecified f-structure with an argument-structure and its semantic role (Bresnan 1996). At this point, arguments are not yet linked to specific grammatical functions and annotated as GF. Relevant semantic information such as tense is also assumed to be available in the input³⁹ as shown in (74).

(74) Underspecified f-structure as the input

$$\left[\begin{array}{l} \text{PRED see}(x, y) \\ \text{GF1["John"]}x \\ \text{GF2["Yoko"]}y \\ \text{TNS 'pres'} \end{array} \right]$$

The grammar GEN(erator) generates a pair of specified f-structure and the c-structure, connected to each other in OT-LFG (Bresnan 1996, Kuhn 2003). GEN generates some possible f-structure, c-structures, and its projection for some language. Afterwards, the constraint ranking chooses the optimal output.⁴⁰ I discuss GEN and constraints in terms of f-structure (section 4.2), c-structure (section 4.3 and 4.4), and their connection (section 4.3 and 4.4).

³⁸ Whether the GEN produces all possible candidates for any languages or only possible candidates for some language is controversial (see Morimoto, 2001).

³⁹ Additionally, the information structure (i-structure) is set up in the input as semantically relevant information in Choi's work (1996). We discuss the issue in the section 4.

⁴⁰ Classically, there is only one optimal candidate. However, as Müller (2000) pointed out in a discussion of free word order, it is often assumed that there are multiple optional forms.

4.2. Argument Structure and Grammatical Function

This section discusses the specification of the f-structure. An f-structure is given as the GF in the input but is not specified as the subject or the object. The GF is specified by means of linking⁴¹ from the semantic representation (e.g., argument structure and thematic role) at the a-structure in the input. The given a-structure in the input (74) can be linked to the grammatical function in two possible ways:

(75) Two candidates for the given input (74)

Input x = agent, y = patient

Candidate 1 Candidate 2

[SUBJ x]	[SUBJ y]
[OBJ y]	[OBJ x]

Two candidates are available in the evaluation by constraint ranking. We need some markedness constraints to evaluate which linking from the semantic role to the grammatical function is unmarked.⁴² Harmonic alignment (Prince and Smolensky 2004) provides context sensitive markedness constraints by combining the prominence scales.⁴³

(76) The prominence of the semantic role and grammatical function

SUBJ(ect) > OBJ(ect) (Aissen 1999)

AG(ent) > PAT(ient)

⁴¹ The specification of the f-structure (A- to F-structure alignment) is one of the constraints that Bresnan (1996) proposed:

Bresnan's OT-LFG constraints:

- I- A- to F-structure Alignment
- II- Faithfulness: All and only the attributes of the input lexical heads appear in input
- III- Economy of Expression: don't project
- IV- C- to F-structure Alignment
 - a. Endocentricity Constraints
 - b. Morphological Constraints

⁴² The prominence of the grammatical function and semantic role is also discussed in classical LFG. Bresnan (1994) and Bresnan and Zaenen (1990) define the following syntactic ranking, and hierarchy of semantic roles, respectively:

SUBJ > OBJ > I.OBJ > OBL > ADJ

agent > beneficiary > experiencer/goal > instrument > patient/theme > locative

The linking between a(rgument) structure and f-structure is assigned in two ways (Bresnan & Zaenen 1990). The most prominent semantic role—agent—is assigned as the subject (not objective argument, [-o]), after which the other semantic roles are assigned to the lowest compatible grammatical function on the markedness hierarchy. Choi (1996) defines the non-subject assignment as the constraint CN2: „Non SUBJ functions align reversely with the c-structure according to the functional hierarchy“

⁴³ The prominence is not the ranking per se (Aissen 1999); markedness ranking is decided according to its context. Harmonic alignment is a way of modeling the markedness reversal, such that one feature is more marked in one context than in the other (Aissen 2002).

(77) Harmonic alignment⁴⁴ (Aissen 1999)

SUBJ/AG > SUBJ/PAT

OBJ/PAT > OBJ/AG

The prominence scale (76) shows that a subject is more prominent than an object and an agent more prominent than a patient (The symbol > means “more prominent than”). The two prominence scales can be aligned and the grammatical function and semantic roles are combined and weighted as shown (77). The symbol > means “more harmonic than”. The agentive subject is more harmonic (or unmarked) than the patientive subject. The patientive object is, on the other hand, more harmonic and unmarked than the agentive object. Accordingly, the unmarked f-structure is clearly the candidate 1, which consists of the agentive subject and the patientive object. I assume that the unmarked f-structure (78) is a canonical one in the language.⁴⁵

(78) The canonical f-structure for the given input (74)

f – structure

$$\left[\begin{array}{l} \text{PRED see}(x,y) \\ \text{SUBJ}["\text{John}"]x \\ \text{OBJ}["\text{Yoko}"]y \\ \text{TNS 'pres'} \end{array} \right]$$

Legendre, Reymond, and Smolensky (1993) provide some account for the emergence of marked constructions such as patientive subject or agentive object as the voice alternation. Thereby, the markedness is triggered by the asymmetric prominence between arguments. Each argument is thereby given as the high or low prominence. Either the low-prominence agent or low-prominent patient triggers the passive/anti-passive voice, respectively.

In the next sections, I assume that the f-structure is specified as (78) in the input given (74).

⁴⁴ The harmonic alignment of grammatical function and semantic role can be also formulated as an “avoid constraint” hierarchy. The combination is a reversal of (77) and penalizes the subject as patient more strongly than the subject as agent:

*SUBJ/PAT >> *SUBJ/AG

*OBJ/AG >> *OBJ/PAT

⁴⁵ I follow Lee (2001) and assume that the „unmarked values are crosslinguistically preferred and basic in all grammars“ and that marked values are „crosslinguistically avoided and used by grammars to express contrast“.

4.3. Identification of grammatical function

Now the f-structure is specified but does not provide the linear order of the grammatical function. The word order can be represented as either a hierarchical structure or as a flat structure at the c-structure. Japanese does not show evidence of the VP-node, and word order is represented flat at the c-structure as discussed in Chapter 2. In the OT-LFG account, Japanese is the language in which the economy principle (“don’t project”) outranks the endocentricity principle:

(79) Bresnan (1996)

Endocentricity > don’t project (economy principle): Configurationality

Don’t project (economy principle) > Endocentricity: Non-configurationality

(80) Economy Principle: Don’t project (Bresnan1996)

(81) Endocentricity Constraint (Bresnan1996)

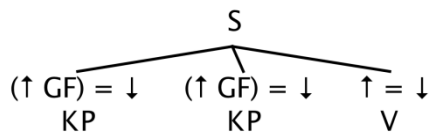
A category and its X’ head correspond to the same f-structure. XP in specifier, complement, or adjunct position corresponds to a function (\uparrow GF) = \downarrow in the local f-structure.

The economy of expression (“don’t project”) is regarded as a variant of the economy principle (e.g. “don’t move”) and it prefers the fewest non-preterminal nodes in the OT-LFG, therefore preferring a flat structure to a hierarchical structure. The markedness principle endocentricity constraint prefers on the other hand the overt coding⁴⁶ of the grammatical function, and therefore a hierarchical syntactical structure. The alignment of the c-structure to the f-structure is thereby realized directly from the node at the c-structure. As for flat structure, I assume that the c-structure is generated by the X-bar theory⁴⁷ with S, and that the noun phrase (here KP) is annotated with the underspecified grammatical function GF: (\uparrow GF) = \downarrow .

⁴⁶ Haspelmath (2006) claims that the term markedness is ambiguous and must be replaced by a more precise formulation. Markedness is used sometimes for high frequency, sometimes for complexity, etc. Overt coding is also regarded as marked as the counter-part of the zero-coded \emptyset .

⁴⁷ I assume that GEN generates some possible c-structure according to X-bar schemata (Bresnan 1996). Sells (2001) explores the property of GEN within X-bar theory (e.g., bar-level, projection, and headness)

(82) Japanese basic phrase structure



Morphology contributes to the projection of specific grammatical function to f-structure. A high-ranked morphological constraint thereby ensures the identification of grammatical function from case marking in Japanese:

(83) Morphological Constraint⁴⁸: Dependent Case (Bresnan, 1996)

Case morphology corresponds to a function $(\uparrow GF) = \downarrow$ in the local f-structure.

However, case markers don't always act as an identifier of grammatical function, as in cases of word order freezing phenomena, in which word order is fixed to determine grammatical function.

(84) 'John likes Yoko'

- a. John=ga Yoko=ga sukida (Strategy: Word Order)
John=NOM Yoko=NOM like.COP.PRES
'John likes Yoko'
- b. ?Yoko=ga John=ga sukida (No Strategy)
Yoko=NOM John=NOM like.COP.PRES
* 'John likes Yoko'
'Yoko likes John'

The sentence (84)b does not encode grammatical function by morphological means, but by syntactic means for the sake of recoverability.⁴⁹ We need for some constraint on

⁴⁸ In this way, economy and markedness constraints compete with each other as well. I assume we need another economy principle for the morphological marking in addition to Bresnan's constraints:

Morphological economy: don't be overt

Japanese has an overt case marker. Therefore, morphological constraints outrank the morphological economy:

Japanese: morphological constraints (markedness) > morphological economy

Aissen (2002) has already discussed as *STRUCc, which states that it penalizes "a value for the morphological category CASE")

noncanonical word order to be restricted in order for sentence (84)a to win the competition among candidate forms:

(85) CANON ORDER:

Identify the grammatical function by the canonical word order!

This constraint penalizes candidates of noncanonical word order, resulting in a penalty for OSV ordering. SOV order thus wins the competition as follows:

(86) Competition between SOV and OSV ordering

Input: like (John, Yoko) John = SUBJ, Yoko = OBJ	Constraint: CANON ORDER
☛ John=ga Yoko=ga (SOV)	
Yoko=ga John=ga (OSV)	*

The constraint CANON ORDER favors SOV order in order to ensure the encoding of the grammatical function if morphology fails at conveying this. Now we need to narrow down in which cases morphology fails to encode grammatical function. To examine purely morphological encoding, we must observe the candidates of noncanonical word order, in which the encoding via word order is excluded:

(87) ‘John sees Yoko’

- a. Yoko=o John=ga mita (Strategy: 2 Case Markings)
Yoko=ACC John=NOM see.PAST
‘John sees Yoko’
- b. Yoko=wa John=ga mita (Strategy: 1 Case Marking)
Yoko=TOP John=NOM see.PAST
‘John sees Yoko’
- c. Yoko=o John=wa mita (Strategy: 1 Case Marking)
Yoko=ACC John=TOP see.PAST
‘John sees Yoko’

⁴⁹ Recoverability means “the ability of a hearer to recover the grammatical relations of argument roles in a sentence” (Lee, 2001).

- d. ? *Yoko=wa John=wa mita* (No strategy)
Yoko=TOP John=TOP see.PAST
 *‘John sees Yoko’
 ‘Yoko sees John’

The sentence (87)a makes use of nominative case *ga* and accusative case *o* in order to identify grammatical function. However, one of these two case markings aligns the argument to the grammatical function successfully, as sentences (87)b and (87)c illustrate. Non-use of the marking is ungrammatical, as in sentence (87)d.

As Aissen (2002) and De Hoop (2008) claim, transitive sentences do not need both arguments to be case-marked. The case-marking system often works enough well to have one instance of the case marking differentiate the arguments. De Hoop (2008) shows that there are two main strategies for the case marking: the identifying strategy and the distinguishing strategy. The identifying strategy encodes specific semantic/pragmatic information such as the volitionality for the ergative case. The distinguishing strategy distinguishes two arguments in a transitive sentence. As for Japanese, the morpheme *wa* is used clearly as part of an identifying strategy. The morpheme *wa* provides a clue as to information structure but not necessarily grammatical function. The topic marker marks both subject and direct object, and cannot distinguish the arguments in transitive sentences. This is more obvious upon comparison with the accusative marker *o*, which marks the object exclusively, and never the subject. The accusative case marking alone differentiates arguments in transitive sentence as shown in sentence (87)c. The nominative case marker *ga* works slightly different from the accusative marker. It exclusively marks subjects in the default case, in which the accusative case marks direct objects. However, the nominative object construction requires a direct object be marked with the nominative case (see the Chapter 2). In this non-default case, the nominative case marks both subject and object. The nominative case does not differentiate arguments and works only as an identifier of information structure.

To avoid the (87)d winning the competition, we need to set a constraint on the differential case marking:

- (88) DISTINCTIVE MORPHOLOGICAL MARKING as DISTINCT-MORPH:
 Differentiate the arguments in the transitive sentence!

The constraint DISTINCT-MORPH requires that at least one of two case markings function as a differentiating marking.

(89) Violation of the constraint DISTINCT-MORPH

Input: see (John, Yoko) John = SUBJ, Yoko = OBJ (SUBJ/OBJ)	Constraint: DISTINCT-MORPH
A. Yoko=o John=ga mita (GA/O ⁵⁰)	
B. Yoko=wa John=ga mita (GA/WA)	
C. Yoko=o John=wa mita (WA/O)	
D. ?Yoko=wa John=wa mita(WA/WA)	*

The sentence (87) corresponds to candidates A, B, C, and D in Table (89). The constraint is fulfilled if one of the case markings is nominative or accusative, which for the predicate *see* is the default case. As such, candidates B and C do not violate the constraint. The use of two differential case markings (candidate A) is not penalized. It does not matter whether one or two differential markings are used. Candidate D violates the constraint DISTINCT-MORPH due to its lack of the differential subject and object case marking. The violation is fatal for candidate D, for neither the morphology nor the word order encodes the grammatical function. The same morpheme marking is unproblematic as long as the candidate has the canonical word order (the sentence (90)e)

(90) ‘John sees Yoko’ (a, b, c, d: the same sentence as (87))

- a. Yoko=o John=ga mita (Strategy: 2 Case Markings)
Yoko=ACC John=NOM see.PAST
‘John sees Yoko’
- b. Yoko=wa John=ga mita (Strategy: 1 Case Marking)
Yoko=TOP John=NOM see.PAST
‘John sees Yoko’
- c. Yoko=o John=wa mita (Strategy: 1 Case Marking)
Yoko=ACC John=TOP see.PAST
‘John sees Yoko’

⁵⁰ The annotation (GA/O) corresponds to the annotation (SUBJ/OBJ). It means that the nominative case *ga* marks the subject and the accusative case *o* marks the object. The annotation is not reflected in the word order.

- d. ? Yoko=wa John=wa mita (No strategy)
 Yoko=TOP John=TOP see.PAST
 *‘John sees Yoko’
 ‘Yoko sees John’
- e. John=wa Yoko=wa mita (Strategy: Word Order)
 John=TOP Yoko=TOP see.PAST
 ‘John sees Yoko’
- f. John=wa Yoko=o mita (Strategy: Word Order & Case Marking)
 John=TOP Yoko=ACC see.PAST
 ‘John sees Yoko’

These sentences show that the one of two strategies, either case marking or word order, is sufficient to determine grammatical function. The use of both strategies is grammatical acceptable but redundant, as shown in the sentence (90)e. In other words, it is unnecessary to fulfill both constraints CANON ORDER and DISTINCT-MORPH; it is not fatal if one of the constraints is violated. It is only fatal if both are violated.

The constraint CANON ORDER and DISTINCT-MORPH are summarized as following:

(91) Ranking: CANON ORDER and DISTINCT-MORPH for the default case

Default: Non-stative verb (e.g., see)			
Candidate		Constraint:	Constraint:
ORDER	MORPHEME SUBJ/OBJ	CANON ORDER	DISTINCT-MORPH
SOV	GA/O		
	GA/WA		
	WA/O		
	WA/WA		*
OSV	GA/O	*	
	GA/WA	*	
	WA/O	*	
	WA/WA	*	*

The SOV word order does not violate grammatical function constraints, as it is possible to detect the grammatical function. In contrast, OSV word order always violates it. As for case marking, the constraint is violated if both a nominative and accusative case marker are missing. This happens in the case of the topic marking WA/WA. The candidate OSV ordering with WA/WA marking violates both constraints and fails to encode

grammatical function. This candidate is exemplified by the unacceptable sentence (90)d. These two constraints depend upon each other and must be regarded as a single constraint. To avoid winning a candidate such as sentence (90)d, these constraints must be unified to penalize the violation of both constraints. By using the local conjunction, these two constraints are unified as one for the compliance of grammatical function encoding:

- (92) The local conjunction (Smolensky 2005)
 A locally conjoined constraint is violated iff both of its conjuncts are violated in a given domain.

The local conjunction can be formulated as:

$C=[C1\&C2]$ D is violated iff both C1 and C2 are violated in a local domain D.

Using the local conjunction, I define the constraint GF-IDENT on the identification of the grammatical function as followings:

- (93) GF-IDENT = [ORDER & MORPH] GF.
 $C = [ORDER \& MORPH] GF$
 $C1 = CANON ORDER$ (as ORDER)
 $C2 = DISTINCT-MORPH$ (as MORPH)
 $D = GF$ (grammatical function identification)

The constraint GF-IDENT is violated only if both constraints ORDER and MORPH are violated. This constraint ensures that grammatical function is encoded in the language. The candidate WA/WA with noncanonical order is violated, for neither word order nor case marker work to identify a subject and an object. We can exclude this type of candidate by ranking the constraint GF-IDENT on the topic of the constraints:

(94) Ranking: GF-IDENT

Candidate		GF-IDENT: [ORDER & MORPH] _{GF}	Constraint X	Constraint Y
word order	morpheme SUBJ/OBJ			
SOV	GA/O			
	GA/WA			
	WA/O			
	WA/WA			
OSV	GA/O			
	GA/WA			
	WA/O			
	WA/WA	*		

The constraint GF-IDENT is ranked highest in the evaluation. I assume the encoding of grammatical function belongs to the core of the grammar and strongly affects acceptability. Its noncompliance will be penalized so strongly that any candidate violating this constraint cannot win the competition even if other high-ranked constraints (here: Constraint X, Constraint Y) are not violated. The next section discusses the ranking of constraints and shows evidence that GF-IDENT is highest ranked.

Now it is necessary discuss the nondefault case with a stative predicate. Stative predicates (e.g., *understand*, *know*, *want*) require the object to be marked with the nominative case. In such cases, a nominative case marking does not differentiate arguments, for it can mark both subject and object. As a result, the subject is not differentiated by morphology. As for the object marking, only the accusative case works as a distinguishing marker, but is not allowed to mark the object in stative predicates. A new constraint is needed to prohibit case markers other than the nominative case to mark the object in the nominative object construction:

(95) FAITH-LEX:
Be faithful to the lexicon!

The constraint FAITH-LEX is violated if the object is not marked by the nominative case. The constraint corresponds to quirky case assignment in LFG. Structural or default case assignment is discussed in next section. The constraint lexicon is competitive to the constraint GF-IDENT:

(96) Violation: Constraint FAITH-LEX

Nominative Object Construction: stative verb			
Candidate		Constraint: GF-IDENT	Constraint: FAITH-LEX
word order	morpheme SUBJ/OBJ		
SOV	GA/O		*
	GA/WA		*
	WA/O		*
	WA/WA		*
	GA/GA		
	WA/GA		
OSV	GA/O		*
	GA/WA	*	*
	WA/O		*
	WA/WA	*	*
	GA/GA	*	
	WA/GA	*	

The constraint GF-IDENT penalizes the OSV ordering with nondifferentiating morphological marking. The nominative case and topic markers do not function as differentiating markers, and the accusative marker survives as the better candidate. The constraint FAITH-LEX penalizes all object markings except nominative case. That result is that OSV ordering violates one of the two highly ranked constraints.⁵¹ That illustrates how impossible it is for an object to appear sentence-initial in the language. As we have seen in the previous chapter, the objective argument must be fronted to optimize information structure. Sentence (97)b shows this problematic case:

- (97) The additional object marker: *no koto* (the same as (28))
- a. Mary no-koto=ga John=ga suki.da
 Mary-GEN-fact=NOM John=NOM like.COP.PRES
 ‘John likes Mary’
- b. ? Mary=ga John=ga suki.da
 Mary=NOM John=NOM like.COP.PRES
 *‘John likes Mary’
 ‘Mary likes John’

⁵¹ The FAITH-LEXICON is lower ranked than the constraint on the identification of grammatical function. I will discuss the evidence for the ranking in the next section.

As previously mentioned in Chapter 2, the phrase *no koto* can function as an additional object affix (or bound clitic), which is mostly optional with the exception of cases such as (97)a. Sentence (97) shows that the additional object marker *no koto* is required if the morphology and word order does not identify the grammatical function. This optionality only occurs if the grammatical function is otherwise encoded as in the following sentences:

(98) The additional object marker *no koto* (the same sentence from (26)a, (26)c, and (27)a, respectively)

- a. John=ga Mary (-no-koto-)=o damashi.ta
 John=NOM Mary (-GEN-fact)=ACC deceive.PAST
 ‘John deceived Mary’
- b. Mary (-no-koto-)=o John=ga damashi.ta
 Mary (-GEN-fact)=ACC John=NOM deceive.PAST
 ‘John deceived Mary’
- c. John=ga Mary (no-koto)=ga suki.da
 John=NOM Mary (-GEN-fact-)=NOM like.COP.PRES
 ‘John likes Mary’

Grammatical function is encoded in all sentences: morphology and word order are used in sentence (98)a. The morphological means are used in the sentence (98)b. Word order is used in (98)c. The optionality of the object marker *no koto* shows its irrelevance. The *no koto* marking is obligatory if neither morphology nor word order encode grammatical function. We now integrate *no koto* marking as object marking as a part of the differential case marking in the constraint GF-IDENT:

(99) *No koto* marking as the differential object marking

The object marking <i>no koto</i> as the candidate (KOTO)		Constraint: GF-IDENT [ORDER & MORPH]GF	Constraint: FAITH-LEX
word order	morpheme: SUBJ/OBJ		
SOV	GA/O		*
	GA/WA		*
	WA/O		*
	WA/WA		*
	GA/GA		
	WA/GA		
	GA/GA + KOTO		
	WA/WA + KOTO		*
	GA/WA + KOTO		*
	WA/GA + KOTO		
OSV	GA/O		*
	GA/WA	*	*
	WA/O		*
	WA/WA	*	*
	GA/GA	*	
	WA/GA	*	
	GA/GA + KOTO		
	WA/WA + KOTO		*
	GA/WA + KOTO		*
	WA/GA + KOTO		

The table shows that the candidate set is aided by the candidate with the *no koto* marking as NO KOTO. As long as *no koto* object marking appears, the constraint GF-IDENT is not violated. Clearly the *no koto* marking works as some sort of last resort to encode grammatical function.

The question is now whether the *no koto* marking has an animacy effect like other differential object markings. Many languages mark objects for case if they bear the higher animacy (Aissen 2002).⁵² The Awtuw, for example, is one of such languages in which the object must be marked with accusative case on the condition that the object is equal to or higher than the subject in the animacy scale (De Hoop 2008).

To examine this animacy effect, it is necessary to test the case with different animacy grades between arguments. Classically, animacy is divided into 3 classical types:

⁵² As for Japanese, case marking is obligatory but often omitted in spoken language. The accusative case *o* is not arbitrarily omitted but can be dropped only if the object bears lower animacy, as Minashima (1993) claims. The written language does not allow such an omission, and requires the noun phrase to be marked with at least either a case-marker or topic/focus-marker. The issue is thoroughly discussed in Lee's work (2007), and I do not discuss it here. However, it is of note that the *no koto* marking is often used with or without case marking in a spoken language. In written Japanese, a *no koto* marking must appear with a case marker or topic/focus marker.

human, animate (e.g., animals), and inanimate (e.g., things). Animacy can be scaled according to its prominence.

(100) Aissen (2002)

Animacy scale: Human > Animate > Inanimate

(101) SUBJ/Human & OBJ/Human

- a. John=wa Yoko=wa suki-da
 John=TOP Yoko=TOP like.COP
 ‘John likes Yoko’
- b. John=wa Yoko-no-koto=wa suki-da
 John=TOP Yoko-GEN-fact=TOP like.COP
 ‘John likes Yoko’
- c. ? Yoko=wa John=wa suki-da
 Yoko=TOP John=TOP like.COP
 *‘John likes Yoko’
 ‘Yoko likes John’
- d. Yoko-no-koto=wa John=wa suki-da
 Yoko-GEN-fact=TOP John=TOP like.COP
 ‘John likes Yoko’

(102) SUBJ/Human & OBJ/Animate

- a. John=wa neko=wa suki-da
 John=TOP cat=TOP like.COP
 ‘John likes the cat’
- b. ⁵³(*) John=wa neko-no-koto=wa suki-da
 John=TOP cat-GEN-fact=TOP like.COP
 ‘John likes the cat’

⁵³ It is difficult to judge the grammaticality of the sentences (102)b and (102)d. The insertion of *no koto* marking is not odd, per se, but rather results in the interpretation that John loves the cat (as opposed to just liking the cat). On the other hand, the sentence would sound strange, if they were to be replaced by the other animals such as bear, rat, snake, etc. I assume that a cat or dog lives near the human being and are personalized as a friend, while bears or rats do not.

- c. ⁵⁴(?)Neko=wa John=wa suki-da
 cat=TOP John=TOP like.COP
 ‘John likes the cat’
 ‘The cat likes John’
- d. (*)Neko-no-koto=wa John=wa suki-da
 cat-GEN-fact=TOP John=TOP like.COP
 ‘John likes the cat’
- (103) SUBJ/Human & OBJ/Inanimate
- a. John=wa sushi=wa suki-da
 John=TOP sushi=TOP like.COP
 ‘John likes sushi’
- b. *John=wa sushi-no-koto=wa suki-da
 John=TOP sushi-GEN-fact=TOP like.COP
 ‘John likes sushi’
- c. Sushi=wa John=wa suki-da
 Sushi=TOP John=TOP like.COP
 ‘John likes sushi’
- d. *Sushi-no-koto=wa John=wa suki-da
 sushi-GEN-fact=TOP John=TOP like.COP
 ‘John likes sushi’

The sentence(101)c, which has a human subject and object, is a candidate that violates GF-IDENT. Grammatical function is encoded by insertion of the *no koto* marking as seen in (101)d. The animacy scale changes the acceptability of the sentence in which none of two strategies identify the grammatical function; the sentence is deemed acceptable if the animacy of the subject is higher than that of the object, as in sentences (102)c and (103)c. Obviously, grammatical function is encoded by the animacy grade. Markedness is illustrated by the harmonic alignment of the animacy scale with grammatical function:

⁵⁴ It is hard to judge which interpretation is preferred.

- (104) Harmonic Alignment⁵⁵ (Aissen(2002)):
 SUBJ/Human > SUBJ/Animate > SUBJ/Inanimate
 OBJ/Inanimate > OBJ/Animate > OBJ/Human

The unmarked value—the human subject and the nonhuman object—functions to encode the grammatical function. Now we must revise the constraint GF-IDENT. As the revision on the constraint [ORDER & MORPH]_{GF}, I propose the extended version [ORDER & MORPH & ANIMACY]_{GF}, which considers the constraint violated if none of the constraints ORDER, MORPH, or ANIMACY are fulfilled in the domain GF.

- (105) EXTENDED GF-IDENT: [ORDER & MORPH & ANIMACY]_{GF}
 Encode the grammatical function by at least one of the following: word order, morphology, or animacy!

The constraint ensures that grammatical function is encoded by either morphological, syntactic, or semantic (i.e., animacy) means⁵⁶. The ANIMACY constraint is not violated if animacy encodes grammatical function.

The *no koto* marking is the last resort for the sentence, in which both arguments have the same animacy, as in sentence (101)d shows. The question now is whether the *no koto* marking is used only for human objects or rather depends on the relative animacy rankings of the arguments.

- (106) SUBJ/animate & OBJ/animate
- a. Neko=wa inu=ga kowai mitai
 cat=TOP dog=NOM afraid seem.PRES
 ‘The cat seems to be afraid of the dog’
- b. Neko=wa inu-no-koto=ga kowai mitai
 cat=TOP dog-GEN-fact=NOM afraid seem.PRES
 ‘The cat seems to be afraid of the dog’

⁵⁵ I assume the following scale of grammatical functions:
 Grammatical Function (Aissen 2002):
 SUBJ(ect) > OBJ(ect)

⁵⁶ Hinds (1982) pointed out that there are five strategies to indicate the grammatical function in Japanese: (1) postpositional particle (2) word order (3) cognitive features (i.e. animacy and semantic role) (4) the „me-first“ principle (i.e. empathy (see Kuno 1987)) (5) world knowledge.

- c. ? Inu=ga neko=wa kowai mitai
 dog=NOM cat=TOP afraid seem.PRES
 *‘The cat seems to be afraid of the dog’
 ‘The dog seems to be afraid of the cat’
- d. Inu-no-koto=ga neko=wa kowai mitai
 dog-GEN-fact=NOM cat=TOP afraid seem.PRES
 ‘The cat seems to be afraid of the dog’
- (107) SUBJ/animate & OBJ/Human
- a. Neko=wa Ningen=ga kowai rashii
 cat=TOP human.being=NOM afraid it-is-said.PRES
 ‘It is said that the cat is afraid of the human being’
- b. Neko=wa Ningen-no-koto=ga kowai rashii
 cat=TOP human.being-GEN-fact=NOM afraid it-is-said.PRES
 ‘It is said that the cat is afraid of the human being’
- c. ? Ningen=ga Neko=wa kowai rashii
 human.being=NOM cat=TOP afraid it-is-said.PRES
 *‘It is said that the cat is afraid of the human being’
 ‘It is said that the human being is afraid of the cat’
- d. Ningen-no-koto=ga Neko=wa kowai rashii
 human.being-GEN-fact=NOM cat=TOP afraid it-is-said.PRES
 ‘It is said that the cat is afraid of the human being’

In these examples, animacy is lower for the subject than for the object, and so does not encode grammatical function. Animacy encoding works only when the animacy of the subject is higher than that of the object, which corresponds to the unmarked animacy construction. The *no koto* marking corresponds clearly to the positive object marking. According to Bossong (1991), the positive marking inside the differential object marking marks the “subject-like objects”. However, the subject-like (i.e., human) object is not allowed being marked with the *no koto* marking as expected, if the subject is inanimate:

- (108) SUBJ/inanimate & OBJ/Human adopted from Matsumoto (2006)
- a. Bara=no kaori=wa ookuno hito=o/*wa
 Rose=GEN odor=TOP many people=ACC/TOP
 hikitsukeru
 attract.PRES
 ‘The odor of roses fascinates many people’
- b. * Bara=no kaori=wa ookuno hito-no-koto=o/wa
 Rose=GEN odor=TOP many people-GEN-fact=ACC/TOP
 hikitsukeru
 attract. PRES
 ‘The odor of roses fascinates many people’
- c. ?Ookuno hito=wa Bara=no kaori=wa
 many people=TOP Rose=GEN odor=TOP
 hikitsukeru
 attract. PRES
 *‘The odor of roses fascinates many people’
 ‘Many people fascinates the odor of roses’
- d. *Ookuno hito-no-koto=o Bara=no kaori=wa
 many people-GEN-fact=ACC Rose=GEN odor=TOP
 hikitsukeru
 attract. PRES
 ‘The odor of roses fascinates many people’

The sentences (108)b and (108)d show that the inanimate subject does not let the human object be marked by the *no koto* marking. Crucially, the morpheme *wa* is not allowed to mark the human object in the canonical and scrambled order. The accusative case marker works as the differential object marking and the topic marker is not allowed to mark the human patient for it is likely that the inanimate agent is identified as the object. In order to realize the information structure like object topicalization, the voice is changed from active to passive as discussed in Watanabe (2000). The most marked pair, inanimate subject and animate object, triggers a voice alternation. That corresponds to the account of

Legendre et al. (1993) that the low-prominence (i.e. inanimate) agent and the high-prominence (i. e. human) patient cause the passive construction.⁵⁷

To sum up: the animacy works as the identifier of grammatical function, if the alignment to the grammatical function is unmarked. The marked animacy combination of arguments triggers a voice alternation. The *no koto* marking is something between them and functions as the last resort for the potential ambiguity of grammatical function identification. To show how the *no koto* marking competes, I set up the strong animacy for both arguments as input, which let it emerge in the language:

- (109) Input: animacy
 John = Human
 Yoko = Human

The next section discusses the realization of the information structure. The new constraints will compete with the more recently introduced constraints [ORDER & MORPH & ANIMACY]_{GF} and FAITH-LEX.

4.4. Realization of Information Structure

We now arrive at the issue of information structure. I assume that information structure is given in the input as part of the relevant semantic information besides a-structure and animacy:

- (110) Input: i-structure
- | | | | |
|---|--|---|--|
| A | TOP {JOHN}
FOC {YOKO}
CONTRAST - | B | TOP {YOKO}
FOC {JOHN}
CONTRAST - |
|---|--|---|--|

The information structure is realized via linear order and morphology in Japanese. The linear order is represented at the c-structure and the morphology at the f-structure. We first discuss the linear order at the c-structure, then the morphology at the f-structure and use of the lexicon.

With regards to c-structure alignment, a constraint may be formulated by using the principle of left or right positioning of the syntactic category KP on the condition that the

⁵⁷ I assume that animate patient is more salient or prominent in discourse, following Malchukov (2008).

Japanese c-structure has a flat structure as its basic phrase structure. To generate the verb at the sentence, it is assumed that the head finality is guaranteed by the high ranking of right head constraints:

- (111) C-structure Constraint⁵⁸
 KP-L >> KP-R
 VP-R >> VP-L

The precedence $KP > VP$ results from these constraint rankings, causing verb-final candidates to win in Japanese. I assume here that KP has the property $(\uparrow GF) = \downarrow$ and $VP \uparrow = \downarrow$ to ensure projection from the c-structure to the f-structure. As for the argument order, the unmarked ordering (SOV) is generated by some highly ranked constraints such as SUBJ(ect)-L(eft) and OBJ(ect)-R(ight):

- (112) F-structure Constraint
 SUBJ-R⁵⁹ >> SUBJ-L (adopted from (Morimoto 2001))
 OBJ-L >> OBJ-R

To simplify the constraints, I combined three constraints in one which prefers the canonical word order SOV.

- (113) F-structure Constraint: CANON-SOV
 The subject aligns right and the object aligns left in a clause, in which the predicate is positioned at the end of the sentence.

The emergence of the marked word order (i.e., OSV) should be triggered by realization of an information structure which prefers the linear order TOPIC-FOCUS. I assume here that the maximal projection KP and VP has the property $(\downarrow PRED FN) \in (\uparrow i DF)$,⁶⁰ so that the discourse function (DF) may be projected from the c-structure to the i-structure. The topic feature is aligned at the left edge and the focus at the right, resulting in the unmarked pragmatic structure, i.e., topic-comment structure. Here, the two rankings are combined in one constraint as TOP-COMT:

⁵⁸ This constraint is adopted from C-structure: Align (Cat, L, α , L) from Sells' work (2001).

⁵⁹ This means that subject aligns with the left edge of the sentence

⁶⁰ Adapted from the work of King (1997). The annotation states that the underspecified Discourse Function (DF) is projected from this node to the i-structure.

(114) I-structure Constraint (adopted from Lee 2001b)

TOP-R >> TOP-L

FOC-L >> FOC-R

As: TOP-COMT: the topic precedes the focus

The constraints CANON-SOV and TOP-COMT relate to the linear order and compete directly, reflecting different interests: satisfying the former requires preserving the canonical word order to encode the grammatical function, whereas the latter selects candidates optimized for realizing the information structure. For example, the input i-structure (110)A prefers the linear order JOHN (topic)-YOKO (focus) so that the unmarked pragmatic structure is realized. The linear order JOHN (subject)-YOKO (object) is the preferred word order from the view of grammatical function as well. Neither TOP-COMT nor CANON-SOV are penalized; the case of input i-structure A results in the canonical word order.

Input (110)B favors the linear order YOKO(topic)-JOHN(focus) for the realization of information structure. This does not, however, encode grammatical function. The i-structure B fulfills only the TOP-COMT and not the CANON-SOV. The ranking of these constraints is thus crucial to decide which selects better candidates.

(115) Ranking: TOP-COMT and CANON-SOV⁶¹

Input: see (John, Yoko) & I-structure B John = SUBJ & FOC Yoko = OBJ & TOP	TOP-COMT	CANON-SOV
A. John=ga Yoko=wa mita (SOV)	*	
◀ B. Yoko=wa John=ga mita (OSV)		*

The ranking TOP-COMT >> CANON-SOV (>> means “is ranked higher than”) lets the better candidate B win the competition. A higher ranked TOP-COMT enables the

⁶¹ The gloss is as follows:

A: John=ga Yoko=wa mita
 John=NOM Yoko=TOP see
 ‘John saw Yoko’
 B: Yoko=wa John=ga mita
 Yoko=TOP John=NOM see
 ‘John saw Yoko’

object with the topic-feature to be fronted. This shows that the realization of pragmatic word order is more strongly preferred than that of grammatical function in Japanese.

Let's examine the ranking of the EXTENDED GF-IDENT.

(116) Ranking: EXTENDED GF-IDENT⁶²

Input: see (John, Yoko) John = SUBJ & FOC.CONTRA Yoko = OBJ & TOP	EXTENDED GF-IDENT	TOP-COMT	CANON-SOV
A. John=wa Yoko=wa mita (SOV)		*	
☛ B.?Yoko=wa John=wa mita (OSV)	*		*

Morphologically, Candidate A and B have the same marking, which is optimal for the information structure. Candidate A is the canonical word order, but is suboptimal at conveying pragmatic structure. Candidate B fulfills, in contrast, the higher ranked topic-comment constraint, yet does not make use of any of the three strategies for identification of grammatical function. The highest constraint EXTENDED GF-IDENT must penalize Candidate B to avoid selecting it as the optimal candidate. The realization of the information structure is, however, still guaranteed by the differential object marker *no koto*, which emerges only if the other three strategies are not to employ.

⁶² The gloss is as follows:

A: John=wa Yoko=wa mita
 John=TOP Yoko=TOP see
 'John saw Yoko'
 B: ?Yoko=wa John=wa mita
 Yoko=TOP John=TOP see
 *'John saw Yoko'
 'Yoko saw John'

(117) *No koto* Marking as the last resort⁶³

Input: see (John, Yoko) John = SUBJ & FOC.CONTRA Yoko = OBJ & TOP	EXTENDED GF-IDENT	TOP-COMT	CANON-SOV
A. John=wa Yoko=wa mita (SOV)		*	
B. Yoko=wa John=wa mita (OSV)	*		*
☛ C. Yoko-no-koto=wa John=wa mita (OSV)			*

Now we come to the issue on the morphological optimization. To examine the morphological competition, I start with the constraint FAITH-LEX, which requires the arguments to be marked with the faithful morpheme to the lexical entry. Stative predicates such as *like* require the nominative to be used to mark the object. That corresponds to the quirky case assignment in LFG, which overrides the default, structural case assignment. The default case requires the object to be marked with the accusative:

- (118) CANON-CASE:
Be faithful to the default case!

Implies
NOM-SUBJ, ACC-OBJ for the non-stative predicate

The constraint CANON-CASE constructs the default case marking. Yet, the nominative construction prefers the object to be marked with the nominative case and the constraint FAITH-LEX, is therefore more highly ranked than the constraint CANON-CASE.

⁶³ The gloss is as follows:

A: John=wa	Yoko=wa	mita
John=TOP	Yoko=TOP	see
'John saw Yoko'		
B: ?Yoko=wa	John=wa	mita
Yoko=TOP	John=TOP	see
* 'John saw Yoko'		
'Yoko saw John'		
C: Yoko-no-koto=wa	John=wa	mita
Yoko-GEN-fact=TOP	John=TOP	see
'Yoko saw John'		

- (119) Ranking: FAITH-LEX and CANON-CASE
FAITH-LEX >> CANON-CASE

I proceed in accordance with Bresnan (2002) and assume that the lexicon is not „the source but the result of syntactic variation, via reranking of the violable universal constraints“. I assume that the default case marking is the result of the markedness constraints:

- (120) Harmonic Alignment:
SUBJ/NOM > SUBJ/ACC > SUBJ/DAT
OBJ/DAT > OBJ/ACC > OBJ/NOM
From the avoidance scale in Woolford (2001):
*dative >> *accusative >> *nominative

The subject is more harmonic with the nominative marking as with the accusative or dative marking, whereas the object is more harmonic with dative or accusative marking than with the nominative marking. The ranking of the default case marking (nominative-accusative) and the quirky case marking (nominative/dative-nominative)⁶⁴ reflect the relative markedness scale (120).

The ranking affects the differential case marking as well. The constraint CANON-CASE stipulates that the subject and object be marked with differential case marking, but the constraint FAITH-LEX prefers being faithful to the quirky-assigned case. As a result, differential case marking is lost: the nominative case does not differentiate the arguments and does not work as the linker to the grammatical function like the topic marker.

Now we need a higher constraint to allow the topic marker in the language:

- (121) Constraint: TOP-WA, CONTRA-WA
TOP-WA: mark the topic feature with the morpheme *wa!*
CONTRA-WA: mark the contrastiveness with morpheme *wa!*

⁶⁴ With regard with the marked construction SUBJ/DAT, there is often the dative subject construction in other languages (such as Korean, German, Spanish, etc)(see Shibatani (1982)). The dative subject is possible to appear in the nominative object construction in Japanese. However, the dative subject sounds strange and appears rarely. It seems to me that the dative case is used only with the topic marker with the some-exclusion reading (contrastiveness). I assume that the constraint OCP is low-ranked for the nominative (but not accusative case) in Japanese and the avoid constraint *SUBJ/DAT is higher ranked than other languages.

The constraint TOP-WA and CONTRA-WA let the topic marker emerge in the language. The topic marker overrides the case feature and the focus feature, and underspecifies the arguments. The constraint TOP-WA enables the more pragmatic marked clitic to appear and is higher ranked than the constraints for the pragmatically unmarked case marking⁶⁵.

(122) The ranking of TOP-WA

TOP-WA >>FAITH-LEX >>CANON-CASE

(123) The ranking of TOP-WA⁶⁶

Input: like (John, Yoko) John = SUBJ & TOP & Human Yoko = OBJ & FOC & Human	TOP-WA	FAITH-LEX	CANON-CASE
A. John=ga Yoko=ga suki-da (GA/GA)	*		*
◀ B. John=wa Yoko=ga suki-da (WA/GA)			**

Non-topic features (i. e. focus) are penalized by the TOP-WA constraint if they are realized via the morpheme *wa*. However, contrastive focus is marked with the topic marker *wa*, and the constraint CONTRA-WA outranks the constraint TOP-WA.

⁶⁵ The case marker identifies the pragmatic feature [-C] [-T]. A case marker assigns the non-contrastive focus in Japanese, but never the topic or the contrastive, as was discussed in Chapter 3.

	TOPIC [+T]	FOCUS [-T]
Contrastive [+C] => some-exclusion	<i>WA</i> <i>Must be clause-initial</i>	<i>WA</i> <i>Must be clause-initial or preverbal</i>
Not-contrastive [-C] => all-exclusion	<i>wa</i> <i>clause-initial</i>	$\neg wa/WA$ <i>Topic < Focus</i>

⁶⁶ The gloss is as follows:

A: John=ga Yoko=ga sukida
John=NOM Yoko=NOM like.COP.PRES
'John likes Yoko'
A: John=wa Yoko=ga sukida
John=TOP Yoko=NOM like.COP.PRES
'John likes Yoko'

(124) The ranking of CONTRA-WA⁶⁷

Input: like (John, Yoko) John = SUBJ & TOP & Human Yoko = OBJ & FOC. CONTRA & Human	CONTRA -WA	TOP- WA	FAITH- LEX	CANON- CASE
A. John=wa Yoko=ga suki-da (WA/GA)	*			**
← B. John=wa Yoko=wa suki-da (WA/WA)		*	*	**

The constraint TOP-WA penalizes candidate B for its use of the morpheme *wa* to mark the focus feature. TOP-WA does not penalize candidate A for the nominative marking *ga*, which does not bear the topic feature. However, the more highly ranked CONTRA-WA prefers that the contrastiveness feature be marked with the topic marker, and so candidate A is penalized. This constraint ranking results in evaluating candidate B as the optimal one.

Now the ranking of the constraint GF-IDENT must be examined.

(125) Ranking of GF-IDENT⁶⁸

Input: like (John, Yoko) John = SUBJ & FOC.CONTRA & Human Yoko = OBJ & TOP & Human	EXTEN DED GF- IDENT	CONTRA- WA	TOP- WA	FAITH -LEX	CANON- CASE
A. Yoko=wa John=wa suki-da (WA/WA)	*		*	*	**
B. Yoko=wa John=ga suki-da (GA/WA)		*		*	*
← C. Yoko-no-koto=wa John=wa suki-da (WA/KOTO + WA)			*	*	**

⁶⁷ The gloss is as follows:

A: John=wa Yoko=ga sukida
John=TOP Yoko=NOM like.COP.PRES
'John likes Yoko'
B: John=wa Yoko=wa sukida
John=TOP Yoko=TOP like.COP.PRES
'John likes Yoko'

⁶⁸ The gloss is as follows:

A: ?Yoko=wa John=wa sukida
Yoko=TOP John=TOP like.COP.PRES
* 'John likes Yoko'
'Yoko likes John'
B: ?Yoko=wa John=ga sukida
Yoko=TOP John=NOM like.COP.PRES
* 'John likes Yoko'
'Yoko likes John'
C Yoko-no-koto=wa John=wa sukida
Yoko-GEN-fact=TOP John=TOP like.COP.PRES
'John likes Yoko'

Candidate A and C have the optimal morphological marking for the input i-structure and violates the fewest relevant constraints. Grammatical function is not, however, identified in any way regarding the candidate A. Morphology and word order fail at capturing grammatical function. Animacy cannot be used due to the symmetric animacy of the arguments. Therefore, candidate A is not grammatical and is penalized by the more highly ranked constraint GF-IDENT. Candidate B realizes the topic with the morpheme *wa* and the focus with *ga*, but its identification of the grammatical function is also problematic. In candidate C, the object affix *no koto* appears as the last resort and differentiates the arguments. It yields the optimal form for the input.

To sum up: This chapter showed that the non-canonical word order is triggered by the realization of pragmatic unmarked structure. That results in failing to encode the grammatical function. The ranking of the morphological marking illustrates that the canonical case marking provides the subject and object differential case marking in the language. The nominative object construction is more marked than the default case and needs some higher constraint to emerge in the language. That results in the loss of the subject and object differential case marking. The higher ranked constraints (e.g., TOP-WA and CONTRA-WA) are needed for the topic marker *wa* to emerge in the language. The highest constraint EXTENDED GF-IDENT ensures that the grammatical function is identified in some way. The positive object marking *no koto* functions as the last resort in the transitive sentence and appears only if all the strategies are unable to identify the grammatical function.

Word Order: EXTENDED GF-IDENT >>TOPIC-COMMENT >>CANONICAL-WORD-ORDER

Morphology: EXTENDED GF-IDENT >> CONTRA-WA >>TOPIC-WA >>FAITH-LEX >>CANONICAL CASE MARKING

5. Conclusion

This thesis discussed the restriction and optimization of word order and morphology to realize grammatical function and information structure in Japanese. At first, the configurationality of Japanese was examined in the LFG framework. Japanese has neither strong evidence for the VP constituent nor for the focus position in the preverbal phrase. As a result, Japanese has neither configurational nor discourse configurational structure at the c-structure, but flat structure. That means that the position of its phrase structure neither projects grammatical function to the c-structure nor discourse function to the i-structure. On the other hand, flat structure is only possible because Japanese morphology encodes grammatical function. Case markers link to specific grammatical functions with some exceptions such as in the word order freezing phenomena. In this case, word order is used for the identification of grammatical function (canonical word order SOV). This morphological failure is not only triggered by the nominative object construction (i.e., double nominative construction) but also by the realization of information structure. The morpheme *wa* is used to identify topicality and contrastiveness but is not specific with regard to the correlation to the grammatical function. Japanese is often considered to be a language which has overt case marking for both arguments, subject and object, and does not have differential case marking. However, this thesis describes the use of differential case marking in Japanese. Especially, the additional object marking *no koto* works as a sort of last resort, appearing only when all three strategies (i.e. word order, case marking, and animacy) are not in play.

To determine the interaction between morphology, word order, and animacy to realize grammatical function and information structure, I used the OT-LFG framework. This analysis shows that Japanese is a language in which information structure is of relevance and allows the direct object to be fronted for the sake of the optimal pragmatic structure. The ranking of the constraints shows that the realization of information structure (TOP-COMT, TOP-WA, CONTRA-WA) is higher ranked than the constraints for the realization of the grammatical function (CANON-SOV, CASE-FAITH-LEX). That is only possible because grammatical function is encoded in multiple ways, whereas information structure is realized in only one way. It is good enough for the grammatical function to be identified by one of three strategies: the two differentiating case markings, word order and animacy. The multiple use of the strategies is redundant. However, the articulation of the

pragmatic structure sometimes pays a price for the grammatical function and it fails to encode the grammatical function. The constraint EXTENDED GF-IDENT is the one which penalizes candidates which encode perfectly the information structure but none of the grammatical function. However, information structure is managed to be realized perfectly in Japanese because the object *no koto* marking is available in the worst case to ensure the encoding of grammatical function in the transitive construction. Furthermore, morphological interaction can be seen in how the case assignments in LFG interact with each other. The default case assignment is the lowest ranked assignment (CANON-CASE) and the quirky case assignment (FEITH-LEX) overranks it. The semantic case assignment (TOP-WA and CONTRA-WA) holds the highest rank.

As a remaining issue, it must be understood that this thesis concerns itself only with Japanese and does not discuss constraints crosslinguistically. The constraints are used to see how relevant information structure is in Japanese and how it interacts with grammatical function encoding. However, it will be closer to the concept of the universal constraint, if more than one language is evaluated on the same but differently weighted constraints. Furthermore, the grammaticalization of the case marker and animacy object affix *no koto* merits discussion. As for the constraints on the word order, I do not discuss the specificity effect in Japanese. Prosody is also of relevance for the realization of information structure in Japanese, but I do not discuss it here due to the difficulty of judgement without empirical data.

6. References

- Aissen, J. 1999. Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory* 17: 673-711.
- Aissen, J. 2002. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21 (3): 435-483.
- Bloom, D. B. 1999. *Case Syncretisms and word order freezing in the Russian Language*. Master thesis, Stanford University.
- Bossong, G. 1991. Differential object marking in Romance and beyond. In *New Analysis in Romance Linguistics: Selected Papers from the XVIII Linguistics Symposium on Romance Languages*, ed. D. Wanner and D. Kibbee, 143–170. Amsterdam/Philadelphia: John Benjamins Publishing.
- Bresnan, J. 1994. Linear order vs. syntactic rank: Evidence from weak crossover. In *Proceedings of the Chicago Linguistic Society 30*. Chicago: Chicago Linguistic Society.
- Bresnan, J. 1996. LFG in an OT Setting: Modelling Competition and Economy. In *LFG Colloquium*. Grenoble, France.
- Bresnan, J. 2000. Optimal syntax. In *Optimality Theory: Phonology, Syntax and Acquisition*, ed. J. Dekkers, F. van der Leeuw, and J. van de Weijer, 334-385. Oxford: Oxford University Press.
- Bresnan, J. 2001. *Lexical-Functional Syntax*. Malden: Blackwell.
- Bresnan, J. 2002. Lexicon in optimality theory. In *The Lexical Basis of Syntactic Processing: Formal, Computational and Experimental Issues*, ed. S. Stevenson and P. Merlo, 39-58. Amsterdam: John Benjamins.
- Bresnan, J. and Zaenen, A. 1990. Deep unaccusativity in LFG. In *Grammatical Relations. A Cross-Theoretical Perspective*, ed. K. Dziwirek, P. Farrell, and E. Mejías-Bikandi, 45–57. Stanford: CSLI Publications.
- Büring, D. 1999. Topic. In *Focus-Linguistic, Cognitive, and Computational Perspectives*, ed. P. Bosh and R. van der Sandt, 142-165. Cambridge, UK: Cambridge University Press.
- Büring, D. 2000. Let's phrase it! Focus, Word Order, and Prosodic Phrasing in German Double Object Constructions. In *Competition in Syntax*, ed. G. Müller and W. Sternefeld. Berlin: Mouton/deGruyter.

- Büring, D. 2007. Semantics, intonation and information structure. In *The Oxford Handbook of Linguistic Interfaces*, ed. G. Ramchand and C. Reiss. Oxford: Oxford University Press.
- Butt, M. and King, T. H. 1996. Structural topic and focus without movement. In *Proceedings of the First LFG Conference*. Grenoble, France.
- Butt, M. and King, T. H. 1998. Interfacing Phonology with LFG. In *Proceedings of the LFG98 Conference*. Brisbane, Australia.
- Butt, M. and King, T. H. 2000. Null Elements in Discourse Structure. In *Papers from the NULLS Seminar*. Delhi: Motilal Banarasidas.
- Butt, M. and King, T. H. 2004a. Case Systems: Beyond Structural Distinctions. In *New Perspectives on Case Theory*, ed. E. Brandner and M. Zinsmeister, 53–87. Stanford: CSLI Publications.
- Butt, M. and King, T. H. 2004b. The status of case. In *Clause Structure in South Asian Language*, ed. V. Dayal and A. Mahajan. Dordrecht: Springer.
- Butt M., King, T. H., Niño, M.-E., and Segond, F. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.
- Choi, H. 1996. *Optimizing the structure in context: Scrambling and Information Structure*. Dissertation, Stanford University.
- Crouch, D., Dalrymple, M., Kaplan R., King, T. H., Maxwell, J., and Newman P. 2008. *XLE Documentation*. Palo Alto Research Center (PARC), http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html
- Dalrymple, M. 2001. *Lexical-Functional Grammar*. San Diego: Academic Press.
- De Hoop, H. 2008. Case strategy. *Linguistic Inquiry* 39 (4): 565-587.
- Dik, S. C. 1997. *The Theory of Functional Grammar*. Vol. 1. Berlin: Mouton de Gruyter.
- Erteschik-Shir, N. 2007. *Information Structure*. New York: Oxford University Press.
- Flack, K. 2007. Ambiguity avoidance as contrast preservation: case and word order freezing in Japanese. In *University of Massachusetts Occasional Papers in Linguistics: Papers in Optimality Theory III*. 32, ed. M. O'Keefe, E. Reilly, and A. W. L Bateman, 57–89.
- Frank, A., King, T. H., Kuhn, J., and Maxwell, J. 2001. Optimality theory style constraint ranking in large-scale LFG grammars. In *Formal and Empirical Issues in Optimality Theoretic Syntax*, ed. P. Sells. Stanford: CSLI Publications.

- Fry, J. and Kaufmann, S. 1998. Information packaging in Japanese. In *Proceedings Joint Conference on Formal Grammar, Head-Driven Phrase Structure Grammar, and Categorical Grammar*. Saarbrücken, Germany.
- Fukui, N. 1995. *Theory of Projection in Syntax*. Stanford: CSLI Publications.
- Gunji, T. 1987. *Japanese phrase structure grammar: a unification-based approach*. Dordrecht: Kluwer.
- Guntel, J. K. Topic-Comment Structure. 1998. In *Studies in Syntactic Typology*, ed. M. Hammond, E. Moravcsik, and J. Wirth. Amsterdam: John Benjamins Publishing Company.
- Hale, K. 1980. Remarks on Japanese phrase structure: Comments on the papers in Japanese syntax. In *Theoretical Issues in Japanese Linguistics: MIT Working Papers in Linguistics 2*: 185-203.
- Hale, K. 1983. Walpiri and the grammar of non-configurational language. *Natural Language and Language Theory* 1: 5-47.
- Haspelmath, M. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42 (1): 25-27.
- Haspelmath, M. 2009. Terminology of case. In *The Oxford Handbook of Case*, ed. A. Malchukov and A. Spencer, 505-517. New York: Oxford University Press.
- Heycock, C. 2007. Japanese wa, ga, and Information Structure. In *Handbook of Japanese Linguistics*, ed. S. Miyagawa and M. Saito. New York: Oxford University Press.
- Hinds, J. 1973. Some remarks on soo su-. In *Papers in Japanese Linguistics*: 18-30.
- Hinds, J. 1982. Case Marking in Japanese. *Linguistics* 20: 541-557.
- Ishihara, S. 2000a. Stress, Focus, and Scrambling in Japanese. In *MIT Working Papers in Linguistics* 39: 142-175.
- Ishihara, S. 2000b. Scrambling and its interaction with stress and focus. *MIT Working Papers in Linguistics* 38: 95-110.
- Kaiser, L. 1998. Representing the structure-discourse iconicity of the Japanese post-verbal construction. In *Functionalism and Formalism in Linguistics*, ed. M. Darnell, E. Moravcsik, and M. Noonan. Amsterdam: Benjamins.
- Kazumitsu, C. 1983. Process on the understanding of the Japanese sentence: the interaction of the understanding strategy. *Shinrigaku-Kenkyuu* 54 (4): 250-256.
- King, T. H. 1995. *Configuring Topic and Focus in Russian*. Stanford: CSLI Publications.
- King, T. H. 1997. Focus domains and information-structure. In *Proceedings of the LFG97 Conference*, CSLI Publications.

- Kishimoto, H. 2004. Transitivity of ergative case-marking predicates in Japanese. *Studies in Language* 28 (1): 105–136.
- Kishimoto, H. 2010. The Semantic Basis of Dative Case Marking in Japanese. In *Kobe Papers in Linguistics* 7: 19-39.
- Koizumi, M. 2008. Nominative Object. In *the Oxford Handbook of Japanese linguistics*, ed. S. Miyagawa and M. Saito. Oxford: Oxford University Press.
- Komagata, N. 2001. Entangled information structure: analysis of complex sentence structure. In *Proceeding of 13th European Summer School in Logic, Language and Information (ESSLLI): Information Structure, Discourse Structure and Discourse Semantics*. Helsinki, Finland.
- Kuhn, J. 2000. Faithfulness violations and bidirectional optimization. *Proceedings of the LFG 2000*, CSLI Publications.
- Kuhn, J. 2003. Optimality-Theoretic Syntax. A Declarative Approach. Stanford: CSLI Publications.
- Kuno, S. 1973. *Structure of the Japanese Language*. Cambridge, MA: MIT Press.
- Kuno, S. 1987. *Functional Syntax: Anaphora, Discourse, and Empathy*. Chicago: University of Chicago Press.
- Kuroda, S. 2005. Focusing on the matter of topic: A study of wa and ga in Japanese. *Journal of East Asian Linguistics* 14: 1-58.
- Lee, H. 2001a. Markedness and Word Order Freezing. In *Formal and Empirical Issues in Optimality-Theoretic Syntax*, ed. P. Sells. Stanford: CSLI Publications.
- Lee, H. 2001b. *Optimization in argument expression and interpretation: A unified approach*. Dissertation, Stanford University.
- Lee, H. 2003. Prominence mismatch and markedness reduction in word order. *Natural Language and Linguistic Theory* 21: 617–680.
- Lee, H. 2007. Case ellipsis at the grammar/pragmatics interface: A formal analysis from a typological perspective. *Journal of Pragmatics* 39: 1465–1481.
- Legendre, G., Raymond, W., and Smolensky P. 1993. An Optimality-Theoretic typology of case and grammatical voice systems. In *Proceedings of the 19th Meeting of the Berkeley Linguistics Society*.
- Li, C. N. and Thompson S. A. 1976. Subject and Topic: A New Typology of Language. In *Subject and Topic*, ed. Chales N. Li. New York: Academic Press.
- Malchukov, A. L. 2008. Animacy and asymmetries in differential case marking. *Lingua* 18, 203-221.

- Matsumoto, Yo. 1996. *Complex Predicates in Japanese*. Stanford: CSLI Publications.
- Matsumoto, Yusuke. 2006. Nihongo-ni okeru museibutsushugo-o tomonau tadoushihyougen: "Vita Sekusuarisu"-o chuushin-ni. (translated: An analysis of Japanese transitive expressions with inanimate agents : a case study of Vita Sexualis). In *Honyakuron Project Ronbunshuu*. Fujisawa: Keio Gijuku Daigaku Shounan Fujisawa Gakkai.
- Mikami, A. 1963. *Nihongo-no-ronri (translated: the logic of the Japanese language)*. Tokyo: Kuroshio.
- Minashima, H. 1993. On the ellipsis of Japanese case particle "o" : a possibility of relevance to animacy and definiteness. *Gengogaku-ronsou*, Special edition: 58-71.
- Miyagawa, S. 1989. *Syntax and Semantics. Structure and Case Marking in Japanese*. San Diego: Academic Press.
- Miyagawa, S. 1997. Against optional scrambling. *Linguistic Inquiry* 28: 1-26.
- Mohanan, T. 1994. *Argument Structure in Hindi*. Stanford: CSLI Publications.
- Molnar, V. 1998. On the Syntax, Phonology, Semantics and Pragmatics of the So-Called "Contrastive Topic" in Hungarian and German. *Acta Linguistica Hungarica* 45 (1-2): 89-166.
- Molnar, V. 2001. Contrast from a contrastive perspective. In *Proceeding of 13th European Summer School in Logic, Language and Information (ESSLLI2001); Information Structure, Discourse Structure and Discourse Semantics*, 99-114. Helsinki, Finland.
- Morimoto, Y. 2001. *The Antisymmetry of Phrase Structure from the Optimality Theoretic Perspective*. Draft. <http://citeseer.ist.psu.edu/494248>.
- Müller, G. 1999. Optimality, Markedness, and Word Order in German. *Linguistics* 37: 777-818.
- Müller, G. 2000. *Elemente der optimalitätstheoretischen Syntax*. Tübingen: Stauffenburg.
- Müller, G. 2002. Free Word Order, Morphological Case, and Sympathy Theory. *Sonderheft der Linguistischen Berichte: Resolving Conflicts in Grammars*, 9-48. Hamburg: Helmut Buske Verlag 2002.
- Muraoka, S., Tamaoka K., and Miyaoka Y. 2005. The Interaction between the Word Order and Animacy in the Understanding. In *Proceeding: The 69th Annual Convention of the Japanese Psychological Association*. Tokyo, Japan.
- Nakau, M. 1973. *Sentential complementation in Japanese*. Tokyo: Kaikakusha.

- Neeleman, A., Titov, E., Van de Koot, H. and Vermeulen, R. 2009. A syntactic typology of topic, focus and contrast. In *Alternatives to Cartography*, ed. J. Van Craenenbroeck. Berlin: Mouton de Gruyter.
- Nishida, M. 2004. A Study of Some Sentence Types in Japanese and of Their Constituents Part Six: Case and Case particles. *Doshisha studies in Language and Culture* 7 (3): 285-364.
- Nordlinger, R. 1998. *Constructive Case. Evidence from Australian Languages*. Stanford: CSLI Publication.
- Ohara, M. 2001. Configurationality in Japanese: How Grammatical Functions are determined. In *Kobe Papers in Linguistics* (3): 74-93.
- Prince, A. and Smolensky, P. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden: Blackwell.
- Reinhart, T. 1981. Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27: 53-94.
- Rooth, M. 1992. A theory of focus interpretation. *Natural Language Semantics* 1: 75–116.
- Rooth, M. Focus. 1996. In *The Handbook of Contemporary Semantic Theory*, ed. L. Shalom, 271–297. Oxford: Blackwell.
- Saito, M. 1985. *Some Asymmetries in Japanese and Their Theoretical Implications*. Dissertation, MIT.
- Saito, M. 1989. Scrambling as semantically vacuous A'-movement. In *Alternative conceptions of phrase structure*, ed. M. Baltin and A. Kroch. Chicago: University of Chicago Press.
- Sells, P. 1990. VP in Japanese: Evidence from -te Complements. In *Japanese/Korean Linguistics, Volume 1*, ed. H. Hajime. Stanford: CSLI Publications.
- Sells, P. 1998. *Japanese Postposing (Involves No Movement)*. Presented at the meeting of the Linguistics Association of Great Britain, Lancaster.
- Sells, P. 2001. *Structure, Alignment and Optimality in Swedish*. Stanford: CSLI Publications.
- Sells, P. 2008. *Oblique Case Marking on Core Arguments in Korean and Japanese*. Workshop on Non-Canonical Perspectives on Case. University of Konstanz, Germany.
- Shibatani, M. 1977. Grammatical Relations and Surface Case. *Language* 53 (4): 789-809.
- Shibatani, M. 1982. Japanese Grammar and Universal Grammar. *Lingua* 57: 103-123.
- Shibatani, M. 1990. *The Language of Japan*. Cambridge: Cambridge University Press.

- Shibatani, M. 2001. Non-canonical constructions in Japanese. In *Non-canonical marking of subjects and objects*, ed. A. Y. Aikhenvald, R. M. W. Dixon and M. Onishi. Amsterdam: John Benjamins.
- Smolensky, P. 2005. Optimality in Phonology II: Harmonic Completeness, Local Constraint Conjunction, and Feature-Domain Markedness. In *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, ed. P. Smolensky and G. Legendre, 590-716. Cambridge: MIT Press.
- Takezawa, K. and Whitman J. 1998. *Kaku-to Gojyun-to Tougokouzou. (translated: Case, Word Order and Syntactic Structure)*. Tokyo: Kenkyusha.
- Tomioka, S. 2007. The Japanese existential possession: A case study of pragmatic disambiguation. *Lingua* 117: 881-902.
- Tomioka, S. 2009. Contrastive topics operate on Speech Acts. In *Information Structure from Different Perspectives*, ed. M. Zimmermann and C. Féry Oxford: Oxford University Press.
- Uechi, A. 1998. *An Interface Approach to Topic/Focus Structure*. Dissertation, University of British Columbia.
- Umbach, C. 2001. Contrast and Contrastive Topic. Contrast from a contrastive perspective. In *Proceeding of 13th European Summer School in Logic, Language and Information (ESSLI2001); Information Structure, Discourse Structure and Discourse Semantics*, 175-188. Helsinki, Finland.
- Valludivi, E. 1993. *Information Component*. Dissertation, University of Pennsylvania.
- Valludivi, E. 1995. Structural Properties of Information Packeging in Catalan. In *Discourse Configurational Languages*, ed. E. K. Kiss, 122-152. New York: Oxford University Press.
- Valludivi, E. and Vilkuna, M. 1998. On rheme and kontrast. In *The limits of syntax*, ed. P. W. Culicover and L. McNally, 79-108. San Diego: Academic Press.
- Vance, T. J. 1993. Are Japanese Particles Clitics? *Journal of the Association of Teachers of Japanese* 27 (1): 3-33.
- Vogel, R. 2004. Correspondence in OT Syntax and Minimal Link Effects. In *Minimality Effects in Syntax*, ed. A. Stepanov, G. Fanselow and R. Vogel. Berlin: Mouton de Gruyter.
- Vogel, R. 2006. Weak Function Word Shift. *Linguistics*: 1059-1093.
- Vogel, R. and Steinbach, M. 1998. The Dative-An Oblique Case. *Linguistische Berichte*: 65-90.

- Watanabe, T. 2000. Object Topicalization, Passive, and Information Structure in Japanese. In *Proceedings of the 14th Pacific Asia Conference on Language, Information and Computation*. Tokyo, Japan.
- Woolford, E. 2001. Case Patterns. In *Optimality-Theoretic Syntax*, ed. G. Legendre, J. Grimshaw and S. Vikner. Cambridge: MIT Press.
- Zaenen, A., Maling, J. and Thráinsson, H. 1985. Case and Gramamtical Functions: The Icelandic Passive. *Natural Language and Linguistic Theory* 3 (4): 441–483.
- Zwicky, A. M. and Pullum, G. K. 1983. Cliticization vs. Inflection: English n't. *Language* 59: 502-13.

E r k l ä r u n g

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit selbständig verfasst und gelieferte Datensätze, Zeichnungen, Skizzen und graphische Darstellungen selbständig erstellt habe. Ich habe keine anderen Quellen als die angegebenen benutzt und habe die Stellen der Arbeit, die anderen Werken entnommen sind – einschliesslich verwendeter Tabellen und Abbildungen - in jedem einzelnen Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht.

Bielefeld, 10. November 2010

Kyoko Sugisaki