Gefördert vom

bmb+f

Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie

Verbmobil

Verbundvorhaben

Japanese Particles in an HPSG Grammar

Melanie Siegel

Universität des Saarlandes

Vm

Report 220
April 1998

April 1998

Melanie Siegel

Fachrichtung 8.7 Allgemeine Linguistik
Universität des Saarlandes
Gebude 17.2
D-66041 Saarbrücken

Tel.: (0681) 302-5284 - 5288
Fax: (0681) 302-5284 - 5341
e-mail: siegel@dfki.de

Gehört zum Antragsabschnitt: 2.4.6 Japanische Grammatik

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen II-01 IV 701 R 4 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei der Autorin.
ISSN 0949-6084

Contents

1 Introduction

In Japanese, case markers and other particles occur after the phrases they mark. Particles can have various functions:

- Case particles mark subcategorized verbal arguments.
- Postpositions mark adjuncts and have semantic attributes.
- Topic particles mark topic or topicalized verbal adjuncts.
- *no* marks an attributive nominal adjunct.

A comprehensive investigation of Japanese particle was missing up to now. General implications were set up without the fact that a comprehensive analysis was carried out. [PS94] mention a manuscript that was written by Tomabechi in 1989 that seems not to be available any more. Two kinds of solutions have previously been proposed: (1) The particles are divided into case particles and postpositions. The latter build the heads of their phrases, while the former do not (cf. [Miy86]). (2) All kinds of particles build the head of their phrases (cf. [Gun87]). Both kinds of analyses lead to problems: If postpositions are heads, while case particles are nonheads, a sufficient treatment of those cases where two or three particles occur sequentially is not possible, as we will show. If on the other hand there is no distinction of particles, it is not possible to encode their different behaviour in subcategorization and modification.

We offer a lexicalist treatment of the problem. Instead of assuming different phrase structure rules we state a type hierarchy of Japanese particles. This makes a uniform treatment of phrase structure as well as a differentiation of subcategorization patterns possible.

Our analysis is based on a large amount of dialogue data. 800 Japanese dialogues concerning appointment scheduling were collected and transcribed in the Verbmobil project, which deals with English, German and Japanese machine translation[1]. These are the basis for the development of the Japanese syntax in Verbmobil. The HPSG[2] syntax that is described in this paper is used for the deep analysis of Japanese dialogues in the project. We clarify the questions which common characteristics and differences between the individual particles exist. A classification in categories was carried out. After that a model hierarchy could be set up for an HPSG grammar. The simple distinction into case particles and postpositions was proved to be as not sufficient.

2 The Japanese Particles

2.1 *ga*

In most cases the *ga*-marked noun phrase is the subject of the sentence:

(1)	naN	nichi	ga	yoroshii	deshou	ka
	which	day	GA	good	COP	QUE

(Which day would suit you?)

However, this is not always the case. Notably stative verbs subcategorize for *ga*-marked objects. An example is the stative verb *dekimasu*:

(2)	kanojo	ga	oyogi	ga	dekimasu
	she	GA	swimming	GA	can

(She can swim.)

These and other cases are sometimes called 'double-subject constructions' in the literature. But these *ga*-marked noun phrases do not behave like subjects. They

[1] see [Wah97] for further information
[2] [PS94]

are not subject to restrictions on subject honorification or reflexive binding. This can be shown by the following example:

(3)

gogo	no	hou	ga	yukkuri	hanashi	ga	dekimasu
afternoon	NO	side	GA	at ease	talking	GA	can
ne							
SAP							

(We can talk at ease in the afternoon.)

hanashi does not meet the semantic restriction [+animate] stated by the verb *dekimasu* for its subject. Nor is it constrained by subject honorification or subject binding of *jibun* in the following variants:

(4)

watashi	ga	yukkuri	hanashi	ga	dekite-orimasu
I	GA	at ease	talking	GA	can-HON

(I can talk at ease.)

The honorification of *dekite-orimasu* does not refer to *hanashi*, but to watashi.

(5)

jibun	ga	yukkuri	hanashi	ga	dekimasu
self	GA	at ease	talking	GA	can

(? can talk at ease.)

The antecedent of *jibun* in 5) is outside of the sentence.
There are even *ga*-marked adjuncts, as in example 3) and the following:

(6)

itsu	ga	go-tsugou	ga	yoroshii	deshou	ka
when	GA	HON-circumstances	GA	good	COP	QUE

(When does it suit you?)

The first NP-*ga* in these examples (3 and 6) is not a subject. It is not subcategorized for by the verb. It is the interrogative word in 6) that is marked by *ga*. *dekimasu* in 3) subcategorizes for two *ga*-marked NPs, but *gogo no hou ga* can neither be the subject nor the object, as it does not fulfill the semantic restrictions for these. [Kur92] assumes these 'double-subject constructions' to be derived from genitive relations. This means that the following sentence from [Far84] is derived from one with a *no*-marked NP:

(7)

yama	ga	ki	ga	kirei	desu
mountain	GA	tree	GA	pretty	COP

(The mountains: Their trees are pretty.)

(8)

yama	no	ki	ga	kirei	desu
mountain	NO	tree	GA	pretty	COP

(The mountain's trees are pretty.)

But this analysis seems not to be true for example 3), because the following sentence is wrong:

(9)

*gogo	no	hou	no	yukkuri	hanashi	ga	dekimasu
afternoon	NO	side	GA	at ease	talk	GA	can
ne							
SAP							

ga marks a verbal adjunct in this example.

To summarize, *ga* is a case particle that usually marks the sentence subject. Sometimes the object is marked by *ga*. This means that the grammatical function is not alloctated by the case particle, but by the verbal valence. In some cases *ga* can even mark an adjunct.

2.2 *o*

The case particle *o* normally marks the direct object of the sentence:

(10)

Sawada	no	hou	ga	zasshi	no	intabyuu	o
PN	NO	side	GA	journal	NO	interview	O
ukemasu	node						
give	SAP						

(Sawada gives an interview to a journal.)

In contrast to *ga*, no two complements may be marked by *o*. This restriction is called 'Double-o-Constraint' in research literature (see, for example, [Tsu96]:249ff.). Consider the following examples from the Verbmobil corpus:

(11)

koNdou	ni	mo	saido	kakuniN	o	shite	mimasu
PN	NI	too	again	confirmation	O	do	try
keredomo							
SAP							

(I will confirm (it) with Mrs. Kondou again.)

(12)

koNdou	no	sukejuuru	o	kakuniN	itashimasu
PN	NO	plan	O	confirmation	HON-do

(I confirm Mrs Kondou's schedule.)

suru can occur with an *o* marked argument or with an unmarked argument. The argument *kakuniN* is marked with *o* in 11) and not marked in 12). The marking in 12) would not be possible, according to the 'double-o constraint', because there is already an *o*-marked argument in the sentence:

(13)

*koNdou	no	sukejuuru	o	kakuniN	o	itashimasu
PN	NO	schedule	O	confirmation	O	tun

The restriction is not valid for embedded sentences:

(14)

koNdou	keNkyuushitsu	no	hou	de	jitsueN	o
PN	institute	NO	side	DE	presentation	O
suru	to iu	yotei	o	tatete iru	N desu	keredomo
do	COMPL	plan	O	built	COP	SAP

(There is a plan to perform the presentation at Mr Kondou's institute.)

Actually there are some violations of the restriction in the Verbmobil data corpus. Examples are:

(15)

kyou	o-deNwa	shita	no	wa	(P) (h) /eto/	hoN
today	telephone	did	NO	WA		book

o	shuppaN	suru	tame ni	(P)	sono	geNkou	o
O	publication	tun	because		that	manuscript	O

itsu	(P)	kou	issho	ni	uchiawase	o	shitara
when		so	joint	NI	appointment	O	do(Kond)

yoroshii	ka	to iu	koto	de	(P)	o-deNwa
good	QUE	COMPL	NOM	DE		telephone

sashite	itadaita	N desu	keredomo
do	HON	COP	SAP

(This is the reason, why I am calling today: When would it suit you to have a joint discussion of that manuscript?)

(16)

uchi	no	satou	ga	(P) /ano/	gakkaishi	no
we	NO	PN	GA		academic journal	NO

tokushuu	no	shuppitsu	keikaku	o	(P) (h)	koNdou
special edition	NO	article	timetable	O		PN

seNsei	to	uchiawase	o	shitai	to	moushite
Prof.	with	appointment	O	want to do	TO	say

orimashita	keredomo
AUX-Past	SAP

(Our Mr. Satou said that he would like to agree upon an appointment to discuss the timetable for the article in the special edition of the academic journal.)

But these examples were described as not acceptable by Japanese native speakers. They are very complex. In both cases there are pauses between the *o* marked entities. The *o* marked nominal phrases *sono geNkou* and *gakkaishi no tokushuu no shuppitsu keikaku* are not subcategorized by *uchiawase*. The examples become acceptabe if one replaces *o* with *nitsuite* and thus marks the NPs as adjuncts. These exceptions of the 'double-o-constraint' are therefore rare effects of spoken language and shall not be introduced into the grammar.

Object positions with *o*-marking as well as subject positions with *ga*-marking can be saturated only once. There are neither double subjects nor double objects. It will be shown that this restriction is also valid for indirect objects. Found arguments must be assigned a saturated status in the subcategorization frame, so that they cannot be saturated again (as in German and English). The verbs subcategorize for at most one subject, object and indirect object. Only one of these arguments may be marked by *o*, while a subject and an object may both be marked by *ga*. These attributes are determined by the verbal valency.

The *o*-marked argument is not adjacent to the verb[3]. It is possible to reverse NP-*ga* and NP-*o* as well as to insert adjuncts between the arguments and the verb:

[3] For the notion of adjacency see [Gun91]

(17)

ikeN	koukaN	o	shimazu	no	hou	ga	sashite
opinion	exchange	O	PN	NO	side	GA	do

itadakitai	to iu	koto	de	o-deNwa	sasete itadakimashita
HON-want	COMPL	NOM	DE	telephone	do-HON-Past

(I've called today because Mr Shimazu would like to exchange opinions (with you))

(18)

paneru disukasshoN	o	koNdo	okonau	N desu	kedomo
panel discussion	O	next time	perform	COP	SAP

(Next time we will perform a panel discussion)

2.3 *ni*

The particle *ni* can have the function of a case particle as well as that of an adjunct particle modifying the predicate. [SK95] also identify homophoneous *ni* that can mark adjuncts or complements. They use the notion of 'affectedness' to distinguish them. This is however not useful in our domain. [Ono94] suggest to test the possibility of passivization. This is helpful in many cases.

Some verbs subcategorize for a *ni*-marked object, as for example *naru*:

(19)

raigetsu	ni	naru	N desu	ga
next month	NI	become	COP	SAP

(It will become next month.)

As *ga*-marked subjects and *o*-marked objects *ni*-marked objects cannot occur twice in the same clause. The 'double-o constraint' is neither a specific Japanese restriction nor a specific peculiarity of the Japanese direct object. It is based on the wrong assumption that grammatical functions are assigned by case particles.

There are a lot of examples with double NP-*ni*. But these are adjuncts, as in the following one:

(20)

juuji	ni	keNkyuushitsu	no	hou	ni	o-ukagai
10 o'clock	NI	institute	NO	side	NI	come

itashimasu
AUX-HON

(I'll come to your institute at 10 o'clock.)

ni as a modifying particle can be found very often in temporal or locative expressions in the Verbmobil data.

(21)

saNji	ni	kaigi	ga	owarimasu
3 o'clock	NI	meeting	GA	end

(The meeting ends at three o'clock.)

(22)

watakushi	ga	sochira	no	keNkyuushitsu	ni	ukagawanai
I	GA	you	NO	institute	NI	not visit

to	ikenai	to	omoimasu	ga
TO	must not do	TO	think	SAP

(I think I'll have to come to your institute.)

2.4 *de*

de can be a verb modifying particle. It has a temporal, locative or instrumental meaning. The temporal meaning of *de* ist restricted to stative verbs:

(23)

asa	juuji	gurai	kara	juuniji	made	no
morning	10 o'clock	ca.	KARA	12 o'clock	MADE	NO
aida	de	yaritai	to	omou	N desu	keredomo
interval	DE	want to do	TO	think	COP	SAP
ikaga	deshou	ka				
good	COP	QUE				

(I would like to do it between 10 and 12 o'clock in the morning. Would that suit you?)

The locative usage of *de* is non-directional:

(24)

keNkyuushitsu	de	jikkeN	no	jitsueN	o	shitai
institute	DE	experiment	NO	performance	O	want to do
N desu	keredomo					
COP	SAP					

(I would like to perform the experiment in the institute.)

An example for the instrumental usage is:

(25)

basu	de	kimasu
bus	DE	come

(I'll come by bus.)

de ist not described as a case particle in literature on this subject. For example [Tsu96] assigns the particle to the class of postpositions. There are, however, examples in the Verbmobil-corpus that show that *de* can also mark verbal arguments:

(26)

Nouchi	seNsei	de	gozaimasu	ka
PN	Prof.	DE	HON-be	QUE

(Are you Professor Nouchi?)

(27)

ichiji	kara	saNji	gurai	no	aida	de
1 o'clock	KARA	3o'clock	ca.	NO	interval	DE
ikaga	deshou	ka				
good	COP	QUE				

(Would it suit you from one o'clock to three o'clock?)

The verb *irasshaimasu* can take a *de*-marked subject. Constructions with adjective and copula also very often occur with a *de*-marked subject. The verb *aru* can also take a *de*-marked argument, but it is the object in this case. The meaning of such a construction is honorific: sentences with a *de*-marked argument signal subject honorification.

This indicates that *de* should also be treated as an ambiguous particle.

2.5 *e*

e is a non-ambigous particle. It is verb modifying and has a directional function cooccuring with verbs of movement. *e* shares this function with *ni*:

(28)

ku ji	ni/*e	sochira	ni/e	ukagaimasu
9 o'clock	NI/*E	you	NI/E	go

(I'll come to you at 9 o'clock.)

2.6 *kara*, *made*, *naNka*

The postpositions *kara* and *made* mark verb modifying adjuncts. These are - as far as the Verbmobil data is concerned - mainly temporal and locative expressions:

(29)

seNsei	no	o uchi	kara	tooi	node	o hiru
Prof.	NO	HON-home	KARA	far	because	HON-noon
kara	ni	shimashou	ka			
KARA	NI	shall do	QUE			

(Shall we start from noon, because it's far from your home?)

Time periods are realized with *kara ... made*:

(30)

kaigi	ga	asa	no	juuichiji	kara	hiru	no
meeting	GA	a.m.	NO	11 o'clock	KARA	p.m.	NO
ichiji	made	arimasu	keredomo				
1 o'clock	MADE	exist	SAP				

(There is a meeting from 11 a.m. to 1 p.m.)

kara as well as *made* can be complements of *desu*:

(31)

saNji	kara	desu	ka
3 o'clock	KARA	COP	QUE

(Does it start at three?)

kara and *made* are non-ambiguous modifying particles, as *e* is. I will call them 'postpositions'. They subcategorize for nominal phrases. Another particle in this category is *naNka*. This word has one function as an adverb and one as a postposition. The postposition *naNka* marks a verb modifying adjunct. An example is:

(32)

gogo	naNka	o-jikaN	yoroshii	deshou	ka
afternoon	NANKA	HON-time	good	COP	QUE

(Would the time in the afternoon be good for you?)

Further postpositions are *to-shite* and *to-shimashite*.

2.7 *no*

no is a particle that modifies nominal phrases. This is an attributive modification and has a wide range of meanings, as the following examples indicate:[4]

(33)	hoka	no	hi
	another	NO	day

(another day)

(34)	tsugi	no	hi
	next	NO	day

(next day)

(35)	watakushi	no	keNkyuushitsu
	I	NO	institute

(my institute)

(36)	nijuukunichi	no	gozenchuu
	29th.	NO	afternoon

(the afternoon of the 29th.)

(37)	kyouto	daigaku	no	kawamura
	Kyoto	university	NO	PN

(Kawamura of Kyoto university)

[Tsu96]:134ff. assigns *no* to the class of case particles. However, the criteria he sets up to distinguish between case particles and postpositions do not apply to this classification of *no*:

- Tsujimura's postpositions have their own semantic meaning. Case particles have a functional meaning. *no* however has a semantic, namely attributive meaning.
- Tsujimura's postpositions are obligatory in spoken language, case particles are optional. *no* is as obligatory as *kara* and *made*.
- Case particles can - as Tsujimura states - follow postpositions, but postpositions cannot follow case particles. According to this criterion, *no* behaves like a case particle.

no combines qualities of case particles with those of modifying particles (which Tsujimura calls 'postpositions'). This means that a special treatment of this particle is necessary.

The particle *no* subcategorizes for a noun, as the other particles do. It also modifies a noun. This separates it from the other modifying particles. NP-*no* is an adjunct to a nominal phrase. As a result, the analysis of multiple NP-*no* is possible:

[4] See also [TH96]

(38)

seNsei	no	hou	no	daigaku	no	keNkyuushitsu
Prof.	NO	side	NO	university	NO	institute

ni	ukagaeba	ii	N desu	ne
NI	go (COND.)	good	COP	SAP

(It would be good to come to your institute, wouldn't it?)

Besides the function of a particle, the word *no* can also have the function of a nominalizer[5]. In this case, it subcategorizes for a verbal head and builds an NP (and can thus be followed by any particle):

(39)

o-deNwa	sasete	itadaita	no	wa	koNdo	no
HON-telephone	do	HON	NOM	WA	next	NO(PART)

koueNkai	naN desu	kedo
talk	COP	SAP

(I am calling because of the next talk.)

2.8 *wa*

The topic particle *wa* can mark arguments as well as adjuncts. In the case of argument marking it replaces the case particle (see example 40). In the case of adjunct marking it can replace the verb modifying particle (see example 41) or it can occur after it (see example 42):

(40)

gogo	wa	aite	orimasu	node
afternoon	WA	be free	HON-AUX	SAP

(The afternoon is free.)

(41)

nijuuhachinichi	no	getsuyoubi	wa	kaigi	ga	gogo
28th	NO	Monday	WA	meeting	GA	afternoon

ni	haitte	orimasu
NI	be inserted	HON-AUX

(On Monday the 28th there is a meeting in the afternoon.)

(42)

koNgetsuchuu	ni	wa	zehi	o-ai shitai	to	omou
in this month	NI	WA	certainly	want to meet	TO	think

N desu	ga
COP	SAP

(I would certainly like to meet you within the month.)

The main problem in the syntactic analysis of *wa* is to decide whether the topic particle marks an argument or an adjunct, when it occurs without a verb modifying particle. This is difficult because of the optionality of verbal arguments in Japanese. If it marks an argument, it has to be decided which grammatical function this argument has. This problem can often not be solved on the purely syntactic level. Semantic restrictions for verbal arguments are necessary:

[5] See [Nig96] for a detailled description of this function of *no*

(43)	basho	no	hou	wa	dou	shimashou	ka
	place	NO	side	WA	how	shall do	QUE

(How shall we resolve the problem of the place?)

Subject and object of the verb *shimashou* are suppressed in this example. The sentence can be interpreted as having a topic adjunct, but no surface subject and object, when using semantic restrictions for the subject (*agentive*) and the object (*situation*).

2.9 *mo*, *koso*

mo is similar to *wa* in some aspects. It can mark a predicative adjunct and can follow *de* and *ni*. But it can also follow *wa*, an adjective and a sentence with question mark:

(44)	gogo	de	mo	kekkou	desu
	afternoon	DE	MO	good	COP

(It would also be good in the afternoon)

(45)	isogashii	mo	node	...
	busy	MO	because	...

(Because I am busy...)

(46)	dekiru	ka	mo	shiremaseN
	can	QUE	MO	do not know

(I don't know if I can)

koso is another topic particle that can occur after nouns, postpositions or adverbial particles.

2.10 *to*

to fulfills a series of extremely different tasks.

Firstly, *to* marks complement sentences that are subcategorized for by verbs like *omou*, *iu* or *kaku*. These complement sentences are adjacent:

(47)	sochira	ni	ukagaitai	to	omoimasu	node
	you	NI	visit	TO	think	SAP

(I would like to visit you.)

(48)	sore	no	hou	ga	ii	to	omoimasu
	that	NO	side	GA	good	TO	think

(I think that would be better.)

Secondly, there are intransitive verbs that only subcategorize for the *to* argument:

(49)	uchiawase	o	shinai	to	ikemaseN	ne
	meeting	WO	do not	TO	not go	SAP

(We have to meet.)

The complement sentence is an utterance that can be marked with sentence particles:

(50)	sorosoro	uchiawase	o	shiyou	ka	to	omou
	soon	meeting	WO	let's do	QUE	TO	think
	no	desu	ga				
	NOM	COP	SAP				

(I think we should soon arrange a meeting.)

Other verbs subcategorize for a *to* marked object:

(51)	kono	hi	mo	chotto	hito	to	au	yotei
	that	day	too	somewhat	people	TO	meet	plan
	ga	gozaimasu						
	GA	exist						

(That day too, there is a plan to meet some people.)

This object can be optional (as in example 51) or obligatory with verbs like *kuraberu*. The particle behaves like a case particle.

The fourth possibility is that *to* marks an adjunct to a predicate:

(52)	shimizu	seNsei	to	teNjikai	o	go-issho	sasete
	Shimizu	Prof.	TO	exhibition	WO	together	do
	itadaku						
	HON						

(I would like to organize an exhibition with Prof. Shimizu.)

In this case, the adjunct is a PN. But *to* can also mark utterances that are adjuncts to other utterances:

(53)	nijuukunichi	desu	to	kaigi	ga	haitte	orimasu
	29th	COP	TO	meeting	GA	inserted	HON

(If it is the 29th, I have a meeting.)

[Shi98] as well as [IG98] describe the relation as a conditional or temporal relation, and not as a conjunctional one.

to can finally be an NP conjunction (which will not be considered at the moment):

(54)	saNjuunichi	to	saNjuuichinichi	wa	chotto	mou
	30th	TO	31rst	WA	somewhat	already
	yotei	ga	haitte	orimasu		
	plan	GA	inserted	HON		

(On the 30th and 31rst there are already plans inserted.)

2.11 Cooccurrence of particles

The discrimination of particles is motivated by their modificational behaviour and by the fact that a Japanese noun phrase can be modified by more than one particle at a time. We carried out an empirical analysis, based on our dialogue data and a questionary with Japanese native speakers[6]. Table 1 shows the possibilities for

[6]Thanks to Atsuko Shimada and Akira Shintani

cooccurrence of some particles.

left↓/right→	ga	wo	ni	de	e	kara	made	no	wa	mo	naNka
ga	-	-	-	-	-	-	-	-	-	-	-
wo	-	-	-	-	-	-	-	-	-	-	-
ni	+	+	+	+	-	-	-	-	+	+	-
de	+	+	+	+	-	-	-	+	+	+	-
e	+	+	+	+	-	-	-	+	+	+	-
kara	+	+	+	+	-	-	-	+	+	+	-
made	+	+	+	+	-	-	-	+	+	+	-
no	-	-	-	-	-	-	-	-	+	-	-
wa	-	-	-	-	-	-	-	-	-	+	-
mo	-	-	-	-	-	-	-	-	-	-	-
naNka	+	+	+	+	-	-	-	-	+	+	-

Table 1: Cooccurrence of Particles

ga can follow the particles *ni*, *de*, *e*, *kara*, *made* and *no*, but not *o* and *wa*. Concerning *ni* and *de*, it has to be said that they can be followed by *ga* only in their modifying function, but not in their case marking function:

(55)

juuji	ni	ga	ikaga	deshou	ka
10 o'clock	NI	GA	good	COP	SAP

(Would 10 o'clock suit you?)

(56)

*raigetsu	ni	ga	naru
next month	NI	GA	become

The reason is obvious: *ga* in combination with another case marking particle would lead to a conflict concerning case marking.

o behaves like *ga*.

ni can follow *de*, *e*, *kara*, *made* and *no*. It cannot follow case particles. Here is an example for *kara–ni*:

(57)

naNji	gurai	kara	ni	shimasu	ka
what time	about	KARA	NI	do	QUE

(At about what time shall we start?)

de also cannot follow case particles. It can follow verb modifying particles in its case marking function:

(58)

saNji	gurai	kara	de	yoroshii	deshou	ka
3 o'clock	about.	KARA	DE	good	COP	QUE

(Would about 3 o'clock suit you?)

In their modifying function *de* and *ni* can follow postpositions, in their case marking function they can follow all kinds of verb modifying particles. It is even possible to have the cooccurrence of the case particle *de* (respectively *ni*) with its modifying counterpart:

(59)	Tokyo	de	de	ikaga	deshou	ka
	Tokyo	DE	DE	good	COP	QUE

(Would it suit you (to meet in) Tokyo?)

kara as well as *made* and *e* cannot follow any other particles. *no* can follow some modifying particles, such as *de*, *e*, *kara* and *made*, but not *ni*. *wa* cannot follow case particles or *mo*, but all other kinds of (analyzed) particles. *mo* behaves like *wa*, except that it can follow it.

In some case three particles occur in a row, as for example:

(60)	gojigoro	made	ni	wa	odeNwa	sashiagemasu
	about 5 o'clock	MADE	NI	WA	telephone	do
	node					
	SAP					

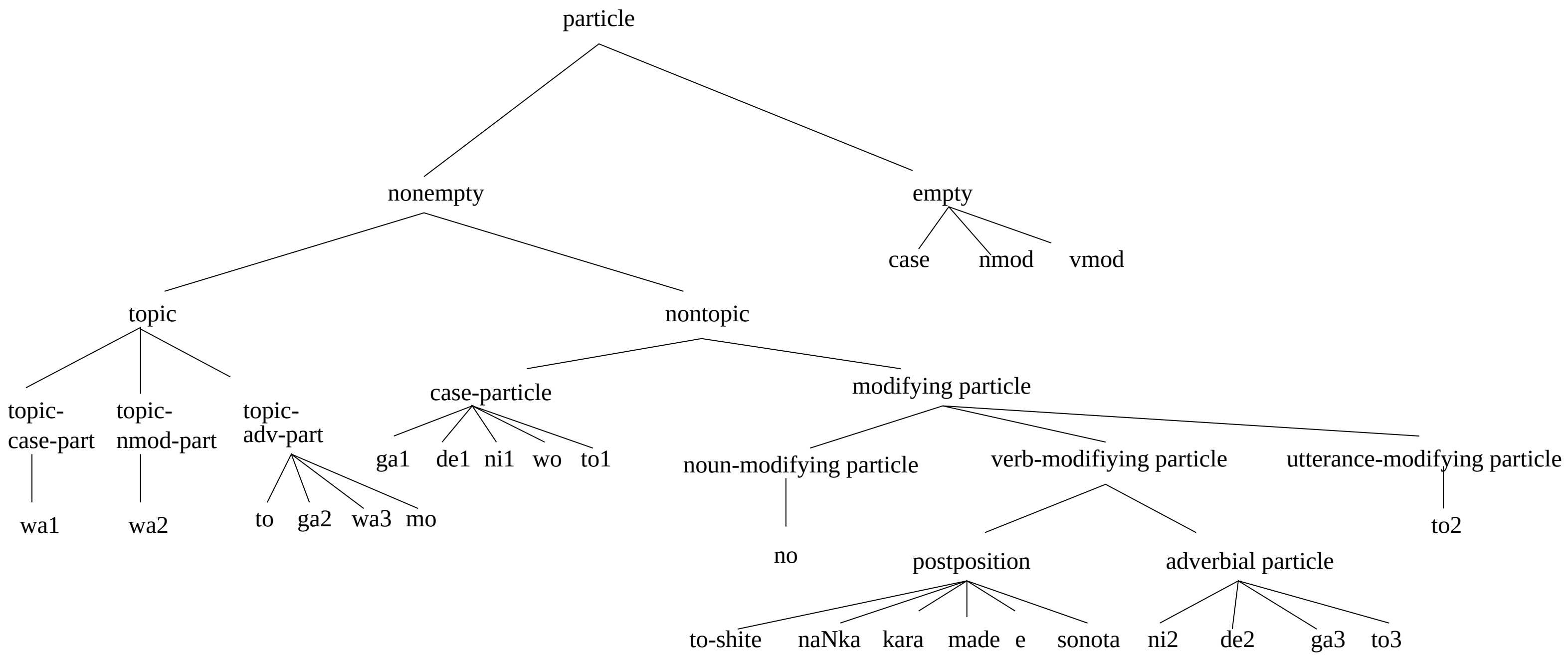
(I will phone you before about 5 o'clock.)

The reason is that *wa* can follow an adverbial particle *ni*. This again can follow a postposition. Another linearization like e.g. *made–wa–ni* or *ni–made–wa* would not be possible.

3 The Type Hierarchy of Japanese Particles

After considering the individual particles, it is now possible to set up a type hierarchy. We shall then feature the individual types of particles.

Firstly, the Japanese particles have to be divided into empty and nonempty particles. Nonempty particles are those with topic function and without topic function. Non-topic particles are further divided into case particles and modifying particles. Case particles receive an entry for case that makes it possible to be subcategorized for by the verbs. Modifying particles get an entry for MOD, specifying the head they can be adjuncts to. The modifying particles separate into noun modifying, verb modifying and utterance modifying ones. Verb modifying particles are postpositions and adverbial particles:


particle
nonempty
empty
case
nmod
vmod
topic
nontopic
topic-case-part
topic-nmod-part
topic-adv-part
case-particle
modifying particle
wa1
wa2
to
ga2
wa3
mo
ga1
de1
ni1
wo
to1
noun-modifying particle
verb-modifiying particle
utterance-modifying particle
no
postposition
adverbial particle
to2
to-shite
naNka
kara
made
e
sonota
ni2
de2
ga3
to3

4 Case Particles

There is no number nor gender agreement between noun phrase and verb. The verbs assign case to the noun phrases. This is marked by the case particles. Therefore these have a syntactic function, but not a semantic one. Different from English the grammatical functions cannot be assigned through positions in the sentence or c-command-relations, since Japanese knows no fixed word position for verbal arguments. Hence, the following variations are possible, for example:

(61)	Hanako	ga	hon	o	kaimasu
	Hanako	GA	book	O	buy

(62)	hon	o	Hanako	ga	kaimasu
	book	O	Hanako	GA	buy

(Hanako buys a book.)

The assignment of the grammatical function is not achieved by the case particle alone but only in connection with the verbal valency. There are verbs that require *ga*-marked objects, while in most cases the *ga*-marked argument is the subject:

(63)	nantoka	yotei	ga	toreru	N desu	ga
	somehow	time	GA	can take	COP	SAP

(Somehow (I) can find some time.)

Japanese is described as a head-final language. [Gun87] therefore assumes only one phrase structure rule: $M \longrightarrow DH$. However, research literature questions whether this also applies to nominal phrases and their case particles. [PS94]:45 also assume Japanese case particles to be markers.

[Miy86] has two arguments for the assumption that the NP is the head in a phrase NP+case particle. He first finds that a distinction between case particles and 'postpositions' is semantically necessary. The reason is that the case particles assign no θ-role to the marked NPs. The second argument concerns the numeral classifiers. They can occur within or outside the NP+case particle (called 'NP' by Miyagawa) which they classify. But they cannot occur outside of an NP+'postposition' (called 'PP' by Miyagawa):

(64)	gakusei	sannin	ga	hon	o	yoNda
	students	3-NK	GA	book	O	read(PAST)

(Exmpl. of [Miy86]: 162):

(65)	*hito	ga	chiisai	mura	kara	futatsu	kita
	people	GA	small	village	from	2	come(PAST)

(66)	hito	ga	futatsu	no	chiisai	mura	kara	kita
	people	GA	2	NO	small	village	from	come(PAST)

(People came from two small villages.)

([Miy86]:157):

(67)

sensei	ga	sannin	kita
teacher	GA	3-CL	come(PAST)

(Three teachers came.)

The restriction that [Miy86]:162 sets up is based on phrase structure:

Definition: X is bijacent to NP, iff:

- X is a sister to NP, or
- X is immediately dominated by a sister of NP.

This restriction for numeral classifiers says that the classifier must be bijancent to the antecedent. Thus, every structure in which the antecedent of the numeral classifier is embedded in a PP is excluded.

Bijacency is however not a sufficient restriction for numeral classifiers, as the following example from [GH98] shows:

(68)

kyoneN	wa	Amerikajin	ga	Nihon	o	3,0000-nin
last year	WA	Americans	GA	Japan	O	30.000 persons
otozureta						
visit(PAST)						

(Last year 30.000 Americans visited Japan.)

It is not possible to set up adequate restrictions on an (exclusively) syntactic base. The phrase-structural distinction between case marked nominal phrases and nominal phrases marked with modifying particles does not further help here. [GH98] show that instead of syntactic restrictions for numeral classificators semantic ones should be used. They use the notions of measurability, coercion, contrastivity and incremental theme in order to explain the phenomena of connection of numeral classifiers and discover two conditions ([GH98]:71):

a *Coercion* Coerced quantification caused by an adverbial measurement.

b *Intervention* Intervention of an adverbially measurable NP in an NP-MP pair.

When both conditions are fulfilled, the sentence is assigned as not accepable.

On the one hand, there are several reasons to distinguish case particles and modifying particles, as has been shown. On the other hand, I doubt whether it is necessary to assume different phrase structures for NP+case particle and NP+modifying particle.

[Yos97]:35 argues that Japanese case particles cannot function as heads, because they can be omitted in spoken language. Ellipsis would be universally seen as a criterion to divide heads and non-heads. However, the ellipsis of heads also often occurs in other languages, as for example in German:

(69)

Wen	hat	Klaus	geküßt?
whom	has	PN	kissed

Maria
PN

(Who did Klaus Kiss? - Maria.)

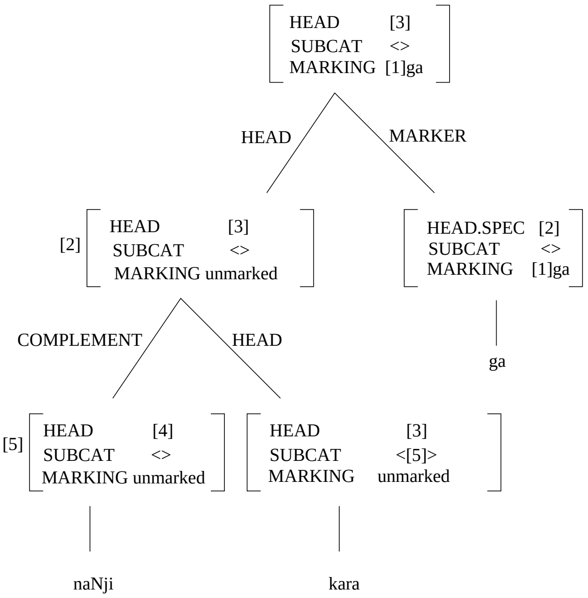
The following example comes from the Verbmobil corpus:

(70)

naNji	kara	ga	yoroshii	desu	ka
what time	from	GA	good	COP	QUE

(At what time would you like to start?)

If one now assumes that the modifying particle *kara* is head of *naNji* as well as of the case particle *ga*, the following results for *naNji kara ga* with the head-marker structure described in [PS94]:[7]



The case particle *ga* would have to allow nouns and modifying particles in SPEC. The latter are however normally adjuncts that modify verbal projections, as the following example shows:

(71)

kochira	kara	seNsei	no	hou	no	keNkyuushitsu
we	from	professor	NO	side	NO	institute
ni	o-ukagai	suru	to iu	katachi	de	yoroshii
NI	HON-visit	do	COMPL	way	DE	good
deshou	ka					
COP	QUE					

(Would it suit you if we come to your institute?)

Therefore the head of *kara* entails the information that it can modify a verb. This information is inherited to the head of the whole phrase by the Head-Feature

[7] The Marking Principle says: *In a headed phrase, the MARKING value is token-identical with that of the MARKER-DAUGHTER if any, and with that of the HEAD-DAUGHTER otherwise[PS94].*

Principle as is to be seen in the tree above.[8] As a result, this is also admitted as an adjunct to a verb, which leads to wrong analyses for the following sentences:

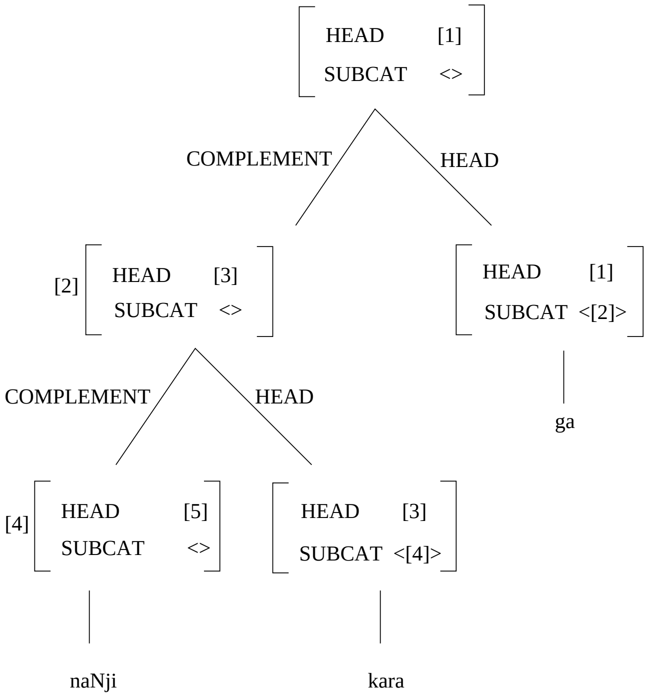
(72)

*naNji	kara	ga	sochira	ga	jikaN	ga	toremasu
what time	from	GA	you	GA	time	GA	can take
ka							
QUE							

(73)

*seminaa	GA	naNji	kara	ga	haitte	irasshaimasu
seminar	GA	what time	from	GA	inserted	AUX-HON
ka						
?						

If, on the other hand, case particles and topic markers are heads, one receives a consistent and correct processing of this kind of example too. This is because the head information [MOD none] is given from the particle *ga* to the head of the phrase *naNji kara ga*. Thus this phrase is not admitted as an adjunct:



[PS94] describe english complementizers as markers. However, a problem results, if *to* is described as a marker. Let's have a look at the following sentence:

(74)

sou	narimasu	to	daibu	saki	ni	natte shimau
so	become	TO	a lot	earlier	NI	become
N desu	ga					
COP	SAP					

[8] The Head Feature Principle says: *The HEAD value of any headed phrase is structure-shared with the HEAD value of the head daughter.[PS94].*

(If it is like this, it will be a lot earlier.)

The complement sentence *sou narimasu* cannot be adjunct to a sentence or a VP. Therefore its head contains the entry [MOD none]:

(75)

*sou	narimasu	daibu	saki	ni	natte shimau	N desu
so	become	a lot	earlier	NI	become	COP
ga						
SAP						

The complement sentence with *to* – on the other hand – can modify a sentence, as example 74) shows. It must therefore have the information [MOD utterance] in his head. The modification could not be realized, if *to* would be marker and *sou narimasu* would be head. Thus, we view *to* as the head of its phrase[9].

Instead of assuming different phrase structure rules, a distinction of the kinds of particles can be based on lexical types. HPSG offers the possibility to define a common type and to set up specifications for the different types of particles.

We assume Japanese to be head-final in this aspect also. All kinds of particles are analysed as heads of their phrases.

The relation between case particle and nominal phrase is a 'Complement-Head' relation. The complement is obligatory and adjacent, as the following examples show:

(76)

*ga
GA

(77)

ie	ga
house	GA

(78)

ookii	ie	ga
large	house	GA

(the large house)

(79)

sono	ookii	ie	ga
that	large	house	GA

(that large house)

(80)

*ie	ookii	ga
house	large	GA

(81)

*ie	sono	ookii	ga
house	that	large	GA

Normally the case particle *ga* marks the subject, the case particle *o* the direct object and the case particle *ni* the indirect object. There are, however, many exceptions. We therefore use predicate-argument-structures instead of a direct assignment of grammatical functions by the particles (and possibly transformations).

[9]See [Mue97] and [Kis95] for an argumentation against analyzing German complementizers as markers, which is similar to ours.

The valency information of the Japanese verbs does not only contain the syntactic category and the semantic restrictions of the subcategorized arguments, but also the case particles they must be annotated with[10].

As in other languages, there are transitive, intransitive and ditransitive verbs in Japanese. However, ditransitive verbs are rare in the processed domain, so none occur in the lexicon up to now. The transitive verbs have to be distinguished depending upon which kind of case particles they require for their subcategorized arguments. The Japanese arguments which are subcategorized for by the verb are optional, unlike the German or English ones. This requires a fundamentally different treatment of subcategorization. We will not adress this point here, however.

- intransitive: NP-ga, e.g.

(82)

watashi	ga	ichinichi	aite-imasu
I	GA	the whole day	be free

(I have the whole day free.)

- ga-o-transitive: NP-ga NP-o, e.g.:

(83)

watashitachi	wa	sanjikan	gurai	jikaN	o	torimasu
we	WA	3 hours	ca.	time	O	take

(We will take about three hours.)

- ga-ga-transitive: NP-ga NP-ga, e.g.

(84)

nantoka	yotei	ga	toreru	N desu	ga
somehow	time	GA	can take	COP	SAP

(I can find some time somehow.)

- aru-transitiv: NP-ni NP-ga or NP-ga NP-ga or NP-ni NP-de or NP-ga NP-de, e.g.

(85)

watakushi	ga	niji	kara	goji	made	kaigi
I	GA	2 o'clock	from	5 o'clock	till	meeting
ga	arimashite					
GA	have					

(I have a meeting from 2 o'clock to five o'clock.)

- naru-intransitiv: NP-ni, e.g.

(86)

itsumo	o sewa	ni	natte	orimasu
always	endeavour	NI	become	AUX(HON)

(Always at your disposal.)

The lexicon entries of case particles get a case entry in the HEAD. Possible values are *ga*, *o*, *ni*, *de* and *to*. They are neither adjuncts nor specifiers and thus get the entries [MOD none] and [SPEC none]. They subcategorize for an adjacent object. This can be a noun.

[10] [Ono94] investigates the particles *ni*, *ga* and *wo* and also states that grammatical functions must be clearly distiguished from surface cases

$$\begin{bmatrix} HEAD & \begin{bmatrix} MAJ & p \\ CASE & case \\ MOD & none \\ SPEC & none \end{bmatrix} \\ SUBCAT & \begin{bmatrix} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun \end{bmatrix} \\ & \end{bmatrix}$$

ga and *o* can, in addition, subcategorize for a verb modifying particle. The case particle *de* can additionally subcategorize for a *ni*-marked phrase, *to* can additionally subcategorize for a VP.

5 Modifying Particles

An essential problem is to find criteria for the distinction of case particles and modifying particles. On the semantic level they can be distinguished in that modifying particles introduce semantics, while case particles have a functional meaning. According to this distinctive feature the particle *no* is a modifying one, because it introduces attributive meaning, as opposed to ([Tsu96]:134), who classifies it as a case particle. Another criterion for distinction that is introduced by [Tsu96]:135 says that modifying particles[11] are obligatory in spoken language, while case particles can be omitted. Case particles are indeed suppressed more often, but there are also cases of suppressed modifying particles. These occur mainly in temporal expressions in our dialogue data:

(87)

soredewa	juuyokka	no	gogo	Ø	niji	Ø
then	14th	NO	afternoon	Ø	2 o'clock	Ø

robii	no	hou	de	o machi	shite	orimasu
lobby	NO	side	DE	HON-wait	do	AUX-HON

(I will then wait in the lobby at 2 o'clock on the 14th.)

Finally Tsujimura gives the criterion that case particles can follow modifying particles while modifying particles cannot follow case particles. This criterion in particular implies that a finer distinction is necessary, as we have shown that it is not that easy. This can be realized with HPSG types. According to this criterion, *no* behaves like a modifying particle, while according to the criterion on meaning, it behaves like a case particle. Our first distincion is thus a functional one: Modifying particles differ from case particles in that their marked entities are not subcategorized for by the verb. Case particles get the head information [CASE case] that controls agreement between verbs and their arguments. Modifying particles do not get this entry. They get the information in MOD that they can become adjuncts to verbs (verb modifying particles) or nouns (the noun modifying particle *no*) and semantic information. They subcategorize for a noun, as all particles do. The modifying particles share the following features in their lexical entries:

[11] He calls them 'postpositions'.

$$\begin{bmatrix} HEAD & \begin{bmatrix} MAJ & p \\ MOD & synsem \\ SPEC & none \end{bmatrix} \\ SUBCAT & \begin{bmatrix} SAT.OBJ & adjacent \end{bmatrix} \\ & \end{bmatrix}$$

5.1 Verb Modifying Particles

The verb modifying particles specify the modification of the verb in SYNSEM/LOC/CAT/HEAD/MOD.

5.1.1 Postpositions

The postpositions are *e*, *kara* and *made*. They modify a (nonauxiliary) verb as an adjunct and subcategorize for a nominal object.

$$\begin{bmatrix} HEAD & \begin{bmatrix} MAJ & p \\ MOD.LOC.CAT.HEAD & nonaux \\ SPEC & none \end{bmatrix} \\ SUBCAT & \begin{bmatrix} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun \end{bmatrix} \\ & \end{bmatrix}$$

5.1.2 Adverbial Particles

[Nig96] treats *ni* and *de* as the infinitive and the gerund form of the copula. To account for this, it has to be clarified what the qualities of an infinitive and a gerundive form, a copula and a verb modifying particle are in our type system. Let us first consider the infinitive form. In our syntax, it has the following peculiarities:

- not honorific concerning addressee
- present tense
- indicative
- possible to use with *n desu* (*jikan ga toreru n desu ka*)
- possible as a relative sentence (*V-ru koto/N*)
- possible as a complement sentence (*V-ru to omou/iu*)
- can be modified by an adverb

This is clearly distinct from the characteristics of *ni*.

The gerundive form, a copula and a verb modifying particle are defined as follows:

- A gerundive form is not finite, it can modify a verbal phrase and be specifier of an auxiliary.
- A copula is a nonauxiliary verb. It subcategorizes for an oblique object, which is an unmarked noun, a postpositional phrase or an adjective. It further subcategorizes for an optional subject, which is marked with *ga*.
- A verb modifying particle is a particle that modifies a nonauxiliar verbal phrase, subcategorizes for an oblique object. I would define adverbial particles as subcategorizing for a noun or a postposition.

The adjunctive form 'de' has both qualities of a gerundive copula and qualities of a particle:

- Subcategorizing for an unmarked noun or a postposition
- Beeing adjunctive to a verbal head
- Its semantic behaviour (see [Nig96])

There are arguments for treating it as a copula:

- Historical derivation (see [Nig96])
- *de arimasu* behaves like *desu*
- The form *deshite* exists

But there is some data that shows different behaviour of *de* and other gerundives. Firstly, it concerns the cooccurrence possibilities of *de* and other particles, compared to gerundive forms and particles:

- de wa - V-te wa
- de mo - V-te mo
- de ga - *V-te ga
- de wo - *V-te wo
- de ni - *V-te ni
- de de - *V-te de
- de no - *V-te no

Secondly, a gerund may modify auxiliaries, e.g. *shite kudasai, shite orimasu*, but *de* may not.

Additionally there is something which distinguishes *de* of a copula: It may not subcategorize for a subject.

A word that is an adjunct to verbs, subcategorizes for an unmarked noun or a phrase with a postposition and is subcategorized for by several particles (see above) fits well into the description of a verb modifying particle.

The adverbial particles *ni*, *de* and *to* subcategorize for a noun or a postposition:

$$\begin{bmatrix} HEAD & \begin{bmatrix} MAJ & p \\ MOD.LOC.CAT.HEAD & nonaux \\ SPEC & none \end{bmatrix} \\ SUBCAT & \begin{bmatrix} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun/postposition \end{bmatrix} \\ & \end{bmatrix}$$

5.2 The Noun Modifying Particle *NO*

The particle *no* modifies a noun phrase and occurs after a noun or a verb modifying particle.

$$\begin{bmatrix} HEAD & \begin{bmatrix} MAJ & p \\ MOD.LOC.CAT.HEAD & noun \\ SPEC & none \end{bmatrix} \\ SUBCAT & \begin{bmatrix} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun/vmod-p \end{bmatrix} \\ & \end{bmatrix}$$

5.3 The Utterance Modifying Particle *to*

to can mark utterances that are adjuncts to other utterances. Thus, HEAD and SUBCAT look as follows:

$$\begin{bmatrix} HEAD & \begin{bmatrix} MAJ & p \\ MOD.LOC.CAT.HEAD & utterance \\ SPEC & none \end{bmatrix} \\ SUBCAT & \begin{bmatrix} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & utterance \end{bmatrix} \\ & \end{bmatrix}$$

5.4 Particles of Topicalization

5.4.1 Topic-*wa*

[Gun91] analyses Japanese topicalization with a trace that introduces a value in SLASH and the 'Binding Feature Principle' that unifies the value of SLASH with a *wa*-marked element[12]. This treatment is similar to the one introduced by [PS94] for the treatment of English topicalization. There as well a trace introduces a SLASH value which is bound by the topicalized element. However, Japanese topicalization is fundamentally different from English one. Firstly, it occurs more frequently. Up to 50% of the sentences are concerned ([Yos97]). Secondly, there are examples where the topic occurs in the middle of the sentence, unlike the English topics that occur sentence-initially. [Yos97] gives the example:

(88)	Bill	ga	Tookyou	e	wa	iku
	PN	GA	Tokyo	E	WA	go

(Bill goes to Tokyo.)

There are also examples in the Verbmobil dialogue corpus:

(89)	raishuuchuu	ni	uchiawase	wa	shitai	N desu	keredomo
	next week	NI	meeting	WA	want to do	COP	SAP

(I would like to hold a meeting in the next week.)

Thirdly, Japanese verbal arguments are optional. Suppressing of verbal arguments could be called more a rule than an exception in spoken language. The SLASH approach would introduce traces in almost every sentence. This, in connection with scrambling and suppressed particles, could not be restricted in a reasonable way. If one follows Gunji interpretation of those cases, where the topic-NP can be interpreted as a noun modifying phrase, a genitive gap has to be assumed. But this leads to assuming a genitive gap for every NP that is not modified. Further, genitive modification can be iterated.

Fourth, two or three occurences of NP-*wa* are possible in one utterance:

(90)	go-yotei	no	hou	wa	raishuu	wa	seNsei	wa
	HON-plan	NO	side	WA	next week	WA	Prof.	WA
	ikaga	deshou	ka					
	good	COP	QUE					

(Concerning your plans: Would next week suit you?)

Thus, we decided to assign topicalized sentences the same syntactic structure as non-topicalized sentences and to resolve the problem on the lexical level.

The topic particle gets three lexical entries, because it is ambiguous. The first one is for the verb modifying variant. Its head is the same as the one of all verb modifying particles. Its subcategorization frame is like the one of adverbial particles.

[12] The Binding Feature Principle says:

> The value of a binding feature of the mother is identical to the union of the values of the binding feature of the daughters minus the category bound in the branching. [Gun91]

$$\left[\begin{array}{ll} HEAD & \left[\begin{array}{ll} MAJ & p \\ MOD.LOC.CAT.HEAD & nonaux \\ SPEC & none \end{array}\right] \\ SUBCAT & \left[\begin{array}{ll} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun/postposition \end{array}\right] \end{array}\right]$$

The second entry is for the case marking variant of *wa*. It gets the same head as the other case marking particles, but the CASE entry is underspecified. It subcategorizes for a nominal complement or a complement with a verb modifying particle.

$$\left[\begin{array}{ll} HEAD & \left[\begin{array}{ll} MAJ & p \\ CASE & case \\ MOD & none \\ SPEC & none \end{array}\right] \\ SUBCAT & \left[\begin{array}{ll} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun/vmod-p \end{array}\right] \end{array}\right]$$

The third one is for the noun modifying variant of the topic particle.

$$\left[\begin{array}{ll} HEAD & \left[\begin{array}{ll} MAJ & p \\ MOD.LOC.CAT.HEAD & noun \\ SPEC & none \end{array}\right] \\ SUBCAT & \left[\begin{array}{ll} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun/vmod-p \end{array}\right] \end{array}\right]$$

5.4.2 Other topic particles

As we have already shown, *mo* is a particle that has the head of a topic-adverbial particle, but a different subcategorization frame as *wa*:

$$\begin{bmatrix} HEAD & \begin{bmatrix} MAJ & p \\ MOD.LOC.CAT.HEAD & nonaux \\ SPEC & none \end{bmatrix} \\ SUBCAT & \begin{bmatrix} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun/vmod-p/ \\ & i-adj/ \\ & utt(modusque) \end{bmatrix} \end{bmatrix}$$

Another topic particle is *koso*:

$$\begin{bmatrix} HEAD & \begin{bmatrix} MAJ & p \\ MOD.LOC.CAT.HEAD & nonaux \\ SPEC & none \end{bmatrix} \\ SUBCAT & \begin{bmatrix} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun/ \\ & postposition/ \\ & adv-p/ \\ & noun \end{bmatrix} \end{bmatrix}$$

6 Omitted Particles

Some particles can be omitted in Japanese spoken language. Here are two examples from the Verbmobil corpus:

(91)

rokugatsu	Ø	juusaNnichi	no	kayoubi	Ø	gogo
June	Ø	13th	NO	Tuesday	Ø	afternoon
kara	wa	ikaga	deshou	ka		
KARA	WA	good	COP	QUE		

(Would the 13th of June suit you?)

(92)

seNsei	Ø	go-tsugou	no	hou	wa	ikaga
Prof.	Ø	HON-circumstances	NO	side	WA	good
deshou	ka					
COP	QUE					

(Would that suit you?)

This phenomenon can be found frequently in connection with pronouns and temporal expressions in the domain of appointment scheduling. [Hin77] assumes that exclusively *wa* can be suppressed. [Yat93] however shows that there are contexts, where *ga*, *o* or even *e* can be omitted. He assigns it as 'phonological deletion'.

[Kur92] analyses omitted *o* particles and explains these with linearization: A particle *o* can only be omitted, when it occurs directly before a verb. [Yat93] however gives examples to prove the opposite. One of these shall be shown here. He assigns it as 'slightly awkward but acceptable':

(93)	dono	gakusei	Ø	ore	ga	nagutta	ka	oboeteru
	which	student	Ø	I	GA	hit	QUE	remember

(Do you remember which student I have hit?)

The Verbmobil data of Japanese dialogues does not contain information about phonological phenomena of pitch. It is therefore not possible at this stage to include this kind of information into our analysis. However, it is peculiar that quite often pauses occur instead of particles. This hints at a phonological phenomenon.

Therefore we propose a phonological-lexical treatment. The empty particle is analyzed as a lexical underspecified particle with empty phonetics. It inherits the features of a particle. It can get the functions of a case particle (as in example 93), of an adverbial particle (as in example 94) or of a noun modifying particle (as in the examples 91 and 92). In any case, it subcategorizes for a noun. It is necessary to distinguish empty and overt particles by a feature called 'PTYPE'. PTYPE can have 'overt' or 'empty' as its value. This is in order to prevent an interpretation with empty particle in, for example, NP+wa as NP+empty-nmod-particle+wa. Therefore, all particles get this additional feature.

(94)	ima	no	tokoro	Ø	gogo	wa	nanimo	yotei
	now	NO	time	Ø	afternoon	WA	no	plan
	ga	haitte	orimaseN	node				
	GA	inserted	HON-NEG	SAP				

(Up to now I have no plans for the afternoon.)

$$\begin{bmatrix} PHON & <> \\ SYNSEM.LOC.CAT & \begin{bmatrix} HEAD & \begin{bmatrix} MAJ & p \\ PTYPE & empty \\ CASE & case \\ MOD & none \\ SPEC & none \end{bmatrix} \\ SUBCAT & \begin{bmatrix} SAT.OBJ & adjacent \\ VAL.OBJ.LOC.CAT.HEAD & noun \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$$
\begin{bmatrix}
PHON & <> \\
SYNSEM.LOC.CAT & \begin{bmatrix}
HEAD & \begin{bmatrix}
MAJ & p \\
PTYPE & empty \\
MOD.LOC.CAT.HEAD & verbal \\
SPEC & none
\end{bmatrix} \\
SUBCAT & \begin{bmatrix}
SAT.OBJ & adjacent \\
VAL.OBJ.LOC.CAT.HEAD & noun
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
PHON & <> \\
SYNSEM.LOC.CAT & \begin{bmatrix}
HEAD & \begin{bmatrix}
MAJ & p \\
PTYPE & empty \\
MOD.LOC.CAT.HEAD & noun \\
SPEC & none
\end{bmatrix} \\
SUBCAT & \begin{bmatrix}
SAT.OBJ & adjacent \\
VAL.OBJ.LOC.CAT.HEAD & noun
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

7 Exhaustivization

One can find several examples with *ga* marked adjuncts in the Verbmobil data. This phenomenon is called 'exhaustivization' in research literature (c.f. [Gun87]), because the meaning of those NP+ga is 'only the NP'. Gunji analyses these syntactically in the same way as he analyzes his 'type-I topicalization'. They build adjuncts that control gaps or reflexives in the sentence. He designs *ga* marked adjuncts without control relations as relying on a very specialized context. Gunjis lexicon entries for exhaustive *ga* are:

a) POS P; PFORM ga; SUBCAT PP [PFORM *pf*; SEM α];
ADJUNCT V [SLASH PP [PFORM *pf*; SEM α]]
where *pf* is not *ga*, *wo*, *ni* or *no*.

b) POS P; PFORM ga; SUBCAT NP [SEM α];
ADJUNCT V [SLASH PP [PFORM *pf*; SEM α]]
where *pf* is *ga*, *wo*, *ni* or *no*.

c) POS P; PFORM ga; SUBCAT NP [SEM α];
ADJUNCT V [REFL PP [SBJ; SEM α]]

However, this treatment leads to the following problems:

1. In all cases, where *ga* marks a constituent that is subcategorized as *ga*-marked by the verb, a second reading is analyzed that contains a *ga* marked adjunct controlling a gap. This is not reasonable. The treatment of the different meaning of *ga* marking and *wa* marking belongs to the semantics and not into the phrase structure.
2. This treatment assumes gaps. We already criticized this in connection with topicalization.
3. The Verbmobil dialogue data contains virtually no reflexives. Therefore, we do not need reflexive control at the moment. However, it contains mostly examples with *ga* marked adjuncts without syntactic control relation to the rest of the sentences.

The examples of the Verbmobil dialogues can be classified into two kinds:

a) The NP describes a temporal entity:

(95)

watakushi	no	hou	no	tsugou	wa	nijuuhachinichi
I	NO	side	NO	circumstances	WA	28th

ga	gogo	ni	kaigi	ga	ikkeN	haitte
GA	afternoon	NI	meeting	GA	at first	inserted

orimasu
HON-AUX

(On our side, there is at first a meeting inserted at the afternoon of the 28th.)

(96)

kochira	wa	getsuyoobi	ga	chotto	sukejuuru
we	WA	Monday	GA	somewhat	schedule

ga	ippai	naN desu	keredomo
GA	full	COP	SAP

(On our side, the schedule is full on Monday.)

b) The NP describes a personal entity:

(97)

watakushi	ga	juuniji	ni	kaigi	ga	owarimasu
I	GA	12 o'clock	NI	meeting	GA	end

(As far as I am concerned, the meeting ends at 12 o'clock.)

All these cases are predicate modifying. To further restrict exhaustivizational interpretations, we introduced selectional restrictions for the marked NP. Only NPs of the sorts 'time' or 'person' are allowed.

8 Conclusion

The syntactic behaviour of Japanese particles has been analyzed using the Verbmobil dialogue data. It has been possible to set up a type hierarchy of Japanese particles. We have therefore adopted a lexical treatment instead of a syntactic treatment based on phrase structure. This is based on the different kinds of modification and subcategorization that occur with the particles. It is necessary to have a finer distinction than just the distinction into case particles and postpositions. The assignment of the grammatical function is done by the verbal valency and not directly by the case particles. The topic particle is ambiguous. Its binding is done by ambiguity and underspecification in the lexicon and not by the Head-Filler Rule as in the HPSG for English ([PS94]). Empty particles are also inserted into the lexicon underspecified, similar to the topic particle. The approach presented here is implemented in the PAGE system [UBB$^+$94]. It is part of the syntactic analysis of Japanese in Verbmobil.

References

[Far84] Ann Kathleen Farmer. *Modularity in Syntax: A Study of Japanese and English.* Massachussetts: MIT Press, 1984.

[GH98] Takao Gunji and Koiti Hasida. Measurement and quantification. In Takao Gunji and Koiti Hasida, editors, *Topics in Constrained-Based Grammar of Japanese*, pages 39–79. Kluwer Academic Publishers, 1998.

[Gun87] Takao Gunji. *Japanese Phrase Structure Grammar.* Dordrecht: Reidel., 1987.

[Gun91] Takao Gunji. An overview of JPSG: A constraint-based descriptive theory for Japanese. In *Proceedings of Japanese Syntactic Processing Workshop.* Duke University, 1991.

[Hin77] John Hinds. Particle deletion in Japanese and Korean. *Linguistic Inquiry*, 8(4):602–604, 1977.

[IG98] Yoshiyuki Igarashi and Takao Gunji. The temporal system in Japanese. In Takao Gunji and Koiti Hasida, editors, *Topics in Constrained-Based Grammar of Japanese*, pages 81–97. Kluwer Academic Publishers, 1998.

[Kis95] Tibor Kiss. *Infinitive Komplementation. Neue Studien zum deutschen Verbum infinitum.* Niemeyer, 1995.

[Kur92] S.-Y. Kuroda. *Japanese Syntax and Semantics. Collected Papers.*, volume 22 of *Studies in Natural Language and Linguistic Theory.* Dordrecht: Kluwer Academic Publishers, 1992.

[Miy86] Shigeru Miyagawa. Predication and numeral quantifiers. In William J. Poser, editor, *Papers from the Second International Workshop on Japanese Syntax*, pages 157–191. CSLI, 1986.

[Mue97] Stefan Mueller. *Spezifikation und Verarbeitung deutscher Syntax in Head-Driven Phrase Structure Grammar.* PhD thesis, Universität des Saarlandes, 1997.

[Nig96] Stephen Nightingale. *An HPSG Account of the Japanese Copula and Related Phenomena.* PhD thesis, University of Edinburgh, 1996.

[Ono94] Kiyoharu Ono. Annularity in the distribution of the case particles *ga*, *o* and *ni* in Japanese. *Theoretical Linguistics*, 20(1):71–93, 1994.

[PS94] C. Pollard and I.A. Sag. *Head-Driven Phrase Structure Grammar.* Chicago: University of Chicago Press., 1994.

[Shi98] Ken-Ichiro Shirai. Japanese conditionals and cost of utterance. In Takao Gunji and Koiti Hasida, editors, *Topics in Constrained-Based Grammar of Japanese*, pages 175–203. Kluwer Academic Publishers, 1998.

[SK95] Kumi Sadakane and Masatoshi Koizumi. On the nature of the "dative" particle *ni* in Japanese. *Linguistics*, 33:5–33, 1995.

[TH96] Hiroshi Tsuda and Yasunari Harada. Semantics and pragmatics of adnominal particle no in Quixote. In Takao Gunji, editor, *Studies in the Universality of Constraint-Based Structure Grammars.* Osaka., 1996.

[Tsu96] Natsuko Tsujimura. *An Introduction to Japanese Linguistics.* Blackwell, Cambridge, 1996.

[UBB+94] Hans Uszkoreit, Rolf Backofen, Stephan Busemann, Abdel Kader Diagne, Elizabeth A. Hinkelman, Walter Kasper, Bernd Kiefer, Hans-Ulrich Krieger, Klaus Netter, Günter Neumann, Stephan Oepen, and Stephen P. Spackman. DISCO—an HPSG-based NLP system and its application for appointment scheduling. In *Proceedings of COLING-94*, pages 436–440, 1994.

[Wah97] Wolfgang Wahlster. Verbmobil. Erkennung, Analyse, Transfer, Generierung und Synthese von Spontansprache. Technical Report 198, DFKI Saarbruecken., 1997.

[Yat93] Shoichi Yatabe. *Scrambling and Japanese Phrase Structure.* PhD thesis, Stanford University., 1993.

[Yos97] Kei Yoshimoto. *Tense and Aspect in Japanese and English.* PhD thesis, Universität Stuttgart, 1997.