

# HPSG Analysis of Japanese

Melanie Siegel

Universität des Saarlandes, Saarbrücken, Germany

**Abstract.** A Japanese HPSG for deep analysis and generation in the Verbmobil system was developed. The focus point of the grammar is the processing of spontaneous Japanese dialogs. Therefore, the description of phenomena of spoken Japanese is central. We present some empirical evaluation of the grammar with Verbmobil corpora.

## 1 Introduction

The fundamental notion of an HPSG is the sign. A lexical sign is a complex feature structure representing information of different linguistic levels of a lexical item. The attribute value matrix (AVM) of a lexical sign in the Japanese Verbmobil grammar can be seen in Figure 2. It is quite similar to the lexical signs of the other HPSG grammars in Verbmobil for English and German, as it contains information about the orthographic realization of the lexical sign in PHON, syntactic information in CAT, semantic information in CONT, information about the lexical status in LEX, nonlocal information in NONLOC, head information (that goes up the tree) in HEAD and information about subcategorization in SUBCAT. As the information contained in HEAD differs in some aspects from the one in the English and German grammars, it needs further explanations that will be given in this article. The structure of the SUBCAT is different, too. This is due to the fundamentally different aspects of subcategorization in Japanese. We will give explanations in Section 4. The CONTEXT needs elaborations for the treatment of Japanese honorification that will be explained in Section 5.

The grammar rules for the Japanese HPSG in Verbmobil are basically quite similar to the other languages. All grammar rules are strictly binary or unary branching. With some exceptions they are head-final.

We will focus on some basic grammatical phenomena of Japanese spoken language and their treatment in the grammar. After that, we will give some evaluations.

## 2 Nominal Constructions

All nouns belong to the super-type *n-lex*. There are nouns that can modify or specify other categories, as will be shown. An ordinary noun does not specify any other category. It can have different honorific forms and it can occur with or without a

particle in spoken language. For example, *raigetsu* (next week) is a non-honorific ordinary noun that cannot occur without particles. *kaigi* (meeting) is a non-honorific ordinary noun that can occur without particles and *kyouju* (professor) is a honorific ordinary noun that can occur without particles. The ordinary noun subcategorizes for an optional specifier, which is a determiner (*kono toki*, this time).

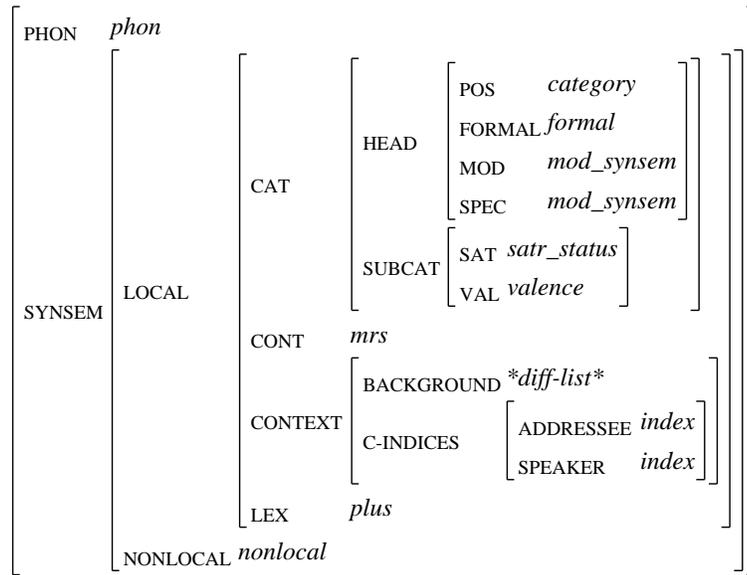


Fig. 1. AVM of a Japanese lexical sign

Names are nouns that are neutral concerning honorification. We have first names, surnames, names of institutions and names of locations. They have different HEADs that are sorted in a type hierarchy of name-heads as can be seen in Figure 2. All names specify a title. First names modify a surname, surnames and institutions do not modify and locational names modify institutions (*aoyama daigaku*, Aoyama University).

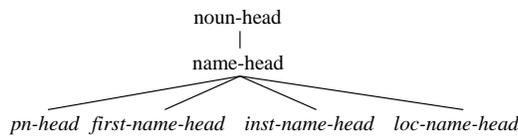


Fig. 2. Type hierarchy of name-heads

Titles are words like *kyouju*, *seNsei*,<sup>1</sup> *saN*, *kuN*, *sama*<sup>2</sup> that subcategorize for a specifier that is a human name, but also institutional titles like *keNkyuushitsu* in *matsumoto keNkyuushitsu* (Matsumoto Institute), *daigaku* (University) or *gaidai* (University). They modify idioms like in *hiratsuka seNsei koNnichiwa* (Hello Professor Hiratsuka) or the main predicate.

Some nouns cannot occur by themselves and need an obligatory argument (see Examples 1 to 4). We call them nominalizations. They take a sentence and nominalize it. Those nominalizing nouns are: *hou*, *koto*, *tame*, *katachi* and *no*. They subcategorize for an obligatory specifier. This specifier can be a verb (... *shita hou ga ii*, it would be better to do), a determiner (*sono hou*, this way) or the particle *no* (*watashi no hou*, my side) in the cases of *hou*, *koto*, *tame* and *katachi* and only a verb in the case of *no*.

*Example 1.*

*kochira no hou de yoji ni owaru koto wa dekimasu kedomo*  
 we NO side DE 4 o'clock NI end NOM WA can SAP  
 (We could end at 4 o'clock.<sup>3</sup>)

*Example 2.*

\**koto wa ii desu*  
 NOM WA good COP

*Example 3.*

*sono tame ni chotto sukejuuru no hou o chousei sasete*  
 that purpose NI somehow schedule NO side WO order do  
*itadakitai to omoimashite*  
 want TO think  
 (For that purpose, I think I want to order my schedule somehow.<sup>4</sup>)

*Example 4.*

\**tame desu*  
 purpose COP

Examples 5 to 8 show some typical examples of temporal expressions in Japanese. Nouns used in temporal expressions are basically ordinary nouns (e.g. *hi*, *shuu*, *jikaN* (day, week, hour)), day-of-week nouns (e.g. *getsuyoubi*, *kayoubi* (Monday, Tuesday)) and temp-numerals (e.g. *gogo*, *asa*, *niji*, *ichigatsu*, *mikka* (afternoon, morning, two o'clock, January, third)). They are subtypes of nouns. Nouns with temporal meaning often occur without particles, as in Example 5. Temp-numerals

<sup>1</sup> Addressing university professors and teachers in general, respectively.

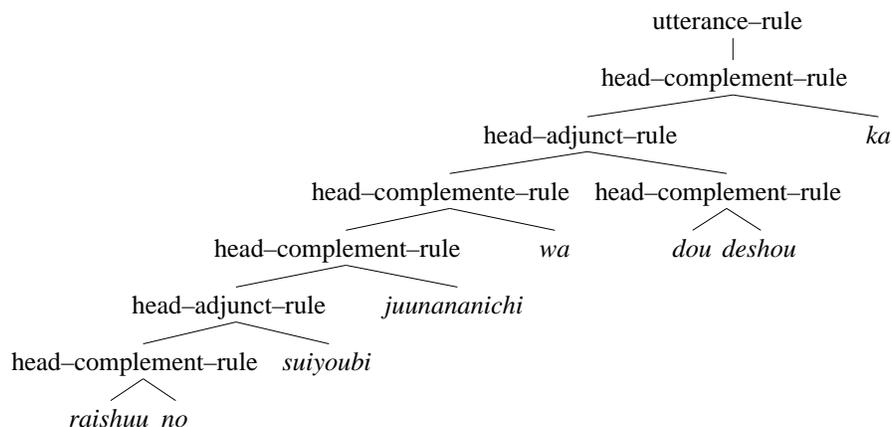
<sup>2</sup> Addressing men and women, with respect to the social distance between speaker and addressee.

<sup>3</sup> From Verbmobil CD16, dialog 010.RMN.

<sup>4</sup> From Verbmobil CD16, dialog 064.RMN.

and day-of-week nouns occur in constructions as in Example 6, where the semantic relation between the two words requires a head-complement structure, while in combinations of ordinary nouns and, e.g. day-of-week nouns there is a head-adjunct relation (see Example 7).

Basically, no special rules are required for the treatment of Japanese temporal expressions, as restrictions are encoded lexically in the subcategorizational and modificational behavior of the lexical types or items. Consider the Example 8 with its chart in Figure 3, where several phrasal types are applied that are used for other constructions as well.



**Fig. 3.** Use of phrasal rules for temporal constructions

*Example 5.*

*rokugatsu juusanichi no kayoubi gogo kara wa ikaga deshou*  
 June 13th NO Tuesday afternoon KARA WA good COP  
*ka*

QUE

(Would Tuesday the 13th of June, in the afternoon, suit you?<sup>5</sup>)

*Example 6.*

*juunananchi no getsuyoubi*  
 17th NO Monday

(Monday the 17th<sup>6</sup>)

*Example 7.*

*raishuu no kayoubi*  
 next week NO Tuesday

(Tuesday of next week<sup>7</sup>)

<sup>5</sup> From Verbmobil CD16, dialog 030.RMN.

<sup>6</sup> From Verbmobil CD25, dialog j009.trl.

<sup>7</sup> From Verbmobil CD17, dialog 349.RMN.

*Example 8.*

*raishuu no suiyoubi juunananichi wa dou desu ka*  
 next week NO Wednesday 17th WA how COP QUE  
 (How would next week Wednesday the 17th be?<sup>8</sup>)

Typically, Japanese nouns are modified by other nouns via the particle *no*, as in Example 7 above. They can also be modified by predicates. This builds the Japanese relative sentence constructions. The noun modification is possible for verbs without addressee honorification, as in Example 9. Another possibility is adjective modification. There are two types of adjectives. The one type directly modifies nouns. It can also be used as a sentence predicative, with predicative inflections. This is analyzed as a subtype of verbs, as will be shown. An example for this kind of noun modification can be seen in Example 10. The other type of adjectives needs the copula form *na* for noun modification<sup>9</sup>. They are subcategorized by *na* and cannot be used in a predicative way (see Example 11).

*Example 9.*

*nakanaka aite iru jikaN ga arimaseN node*  
 more and more free be time GA not exist sentence-part.  
 (There is less and less free time.<sup>10</sup>)

*Example 10.*

*choudo ii jikaN da to omoimasu*  
 just good time COP TO think  
 (I think this is just a good time.<sup>11</sup>)

*Example 11.*

*kirei na hoteru ni tomatte mitai*  
 beautiful COP hotel NI stay want to try  
 (I want to try to stay in a beautiful hotel.<sup>12</sup>)

Numeral classifiers like *futatsu*, *ippoN* (two, one) or *ikkeN* (one) can occur before or behind the phrase they modify. In our data, only PPs with case particles are involved, as in Example 12. The classifiers can also occur by themselves, just as ordinary nouns (Example 13).

*Example 12.*

*jugyou ga futatsu haitte orimashite*  
 class GA two included have  
 (I have two classes.<sup>13</sup>)

<sup>8</sup> From Verbmobil CD19, dialog 667.RMN.

<sup>9</sup> See Nightingale (1996) for *na* as a copula construction.

<sup>10</sup> From Verbmobil CD16, dialog 005.RMN.

<sup>11</sup> From Verbmobil CD27 dialog j034ac.trl.

<sup>12</sup> From Verbmobil CD27 dialog j036ac.trl.

<sup>13</sup> From Verbmobil CD16, dialog 091.RMN.

*Example 13.*  
*mittsu naN desu keredomo*  
 (It's three.<sup>14</sup>)

### 3 The Treatment of Particles in a Type Hierarchy

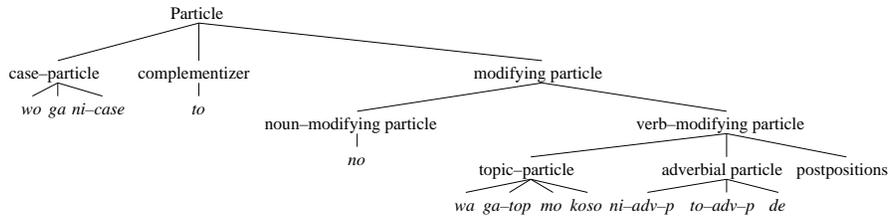
The treatment of particles is essential for the processing of Japanese language for two reasons. The first reason is that these are the words that occur most frequently. The second reason is that particles have various central functions in the Japanese syntax: Case particles mark subcategorized verbal arguments, postpositions mark adjuncts and have semantic attributes, topic particles mark topic adjuncts or topicalized verbal arguments and *no* marks an attributive nominal adjunct. Their treatment is difficult for three reasons: 1) Despite their central position in Japanese syntax, the omission of particles occurs quite often in spoken language. 2) One particle can fulfill more than one function. 3) Particles can co-occur, but not in an arbitrary way. In order to set up a grammar that accounts for a large amount of spoken language, a comprehensive investigation of Japanese particles is thus necessary<sup>15</sup>.

Japanese noun phrases can be modified by more than one particle at a time. There are many examples in our data where two or three particles occur sequentially. On the one hand, this phenomenon must be accounted for in order to attain a correct processing of the data. On the other hand, the discrimination of particles is motivated by their modificational and subcategorizational behavior. We carried out an empirical analysis, based on our dialog data and a questionnaire with Japanese native speakers. Based on this observation, we set up a type hierarchy of Japanese particles to be used in the Verbmobil HPSG grammar for Japanese. Into the class of particles, we include case particles, complementizers, modifying particles and conjunctive particles. We thus assume a common class of the several kinds of particles. One major motivation for the type hierarchy is the observation we made of the co-occurrence of particles. Case particles (*ga*, *wo*, *ni*) are those that attach to verbal arguments. A complementizer marks complement sentences. Modifying particles attach to adjuncts. They are further divided into noun-modifying particles and verb-modifying particles. Verb-modifying particles can be topic particles, adverbial particles, or postpositions. Some particles can have more than one function, as for example *ni* has the function of a case particle and an adverbial particle. Figure 4 shows the type hierarchy of Japanese 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. The assignment of the grammatical function is not achieved by the case particle alone but only in connection with the verbal valence. There are verbs that

<sup>14</sup> From Verbmobil CD19, dialog 792.RMN.

<sup>15</sup> See Siegel (1998) and Siegel (1999) for a more detailed description.



**Fig. 4.** Type hierarchy of Japanese particles. Postpositions include *e*, *naNka*, *sonota*, *tomo*, *kara*, *made*, *soshite*, *nado*, *bakari*, *igai*, *yor*, *toshite*, *toshimashite*, *nitsuite*, *nikaNshite* and *nikakete*

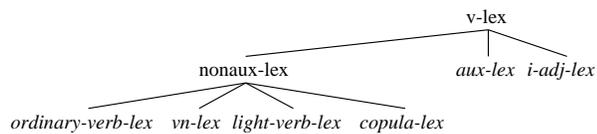
require *ga*-marked objects, while in most cases the *ga*-marked argument is the subject.

Instead of assuming different phrase structure rules for different kinds of particles, a distinction 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. All kinds of particles are analyzed as heads of their phrases. The relation between case particle and nominal phrase is a complement-head relation. The complement is obligatory and adjacent.

## 4 Verbal Constructions

### 4.1 Lexical Hierarchy of Verbs and Adjectives

In Figure 5 the type hierarchy of Japanese verbs can be seen. The first division of verbs is non-auxiliaries vs. auxiliaries vs. adjectives. Non-auxiliaries may be ordinary verbs, verbal nouns, light verbs and copula verbs. There are several subtypes of ordinary verbs and verbal nouns, due to their subcategorizational behavior.



**Fig. 5.** Type hierarchy of v-lex

A fundamental difference between Japanese grammar and English or German grammar is the fact that verbal arguments are frequently omitted. For example, subjects and objects that refer to the speaker are omitted in most cases in spoken language (see Example 14). Additionally, verbal arguments can freely scramble. On the other hand, there exist obligatory and adjacent verbal arguments, as can be seen in the Examples 15 to 18.

*Example 14.*

*hinichi o kimetai*  
 day WO want to decide  
 ((I) want to decide on a day.<sup>16</sup>)

*Example 15.*

*kaigi wa nijikan gurai desu*  
 Meeting WA 2 hours about COP  
 (The meeting is about two hours.)

*Example 16. \*kaigi wa desu.*

*Example 17. \*desu.*

*Example 18. \*nijikaN gurai kaigi wa desu.*

To account for this, our subcategorization contains the attributes SAT and VAL. In SAT it is noted, whether a verbal argument is already saturated (such that it cannot be saturated again), optional or adjacent. VAL contains the agreement information for the verbal argument. The Japanese head-complement schema is different from standard HPSG in the treatment of subcategorization. It must account for optional and scrambable arguments as well as for obligatory and adjacent arguments (see Figure 6).

Adjacency must be checked in every rule that combines heads and arguments or adjuncts. This is stated in the principle of adjacency, formulated as follows:

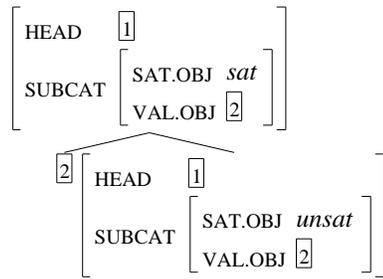
In a headed phrase, the SUBCAT.SAT value of the non-head daughter must not contain any adjacent arguments. In a head-complement structure, the SUBCAT.SAT value of the head daughter must not contain any adjacent arguments besides the non-head daughter. In a head-adjunct structure, the SUBCAT.SAT value of the head daughter must not contain any adjacent arguments.

## 4.2 Verbal Nouns and Light Verbs

A special treatment is needed for Japanese verbal noun–light verb constructions. In these cases, a word that combines qualities of a noun with those of a verb occurs in a construction with a verb that has no restrictions for a subject and only marginal semantic information. In 19 to 21 there are some examples from the corpus. The verbal noun does not inflect. Though, it subcategorizes, can be intransitive, transitive and ditransitive and gives sortal restrictions for its arguments. It is adjacent and obligatory to the light verb<sup>17</sup>. The predicate is formed by the complex. Research literature

<sup>16</sup> From Verbmobil CD25 dialog j006.trl

<sup>17</sup> Dubinsky (1997) explains the atypical syntactic behavior of verbal nouns.



**Fig. 6.** The Japanese head-complement schema

(e.g. Grimshaw and Mester, 1988) talks about so-called “argument transfer”, where the arguments of the verbal noun are transferred to the light verb. Our analysis is based on the viewpoint that the verbal noun—light verb complex is sub-syntactic. It needs a special rule that allows to combine the information from both components. We call it vn-light rule and use the specifier position for the verbal noun. See in Figure 7 how the information is combined. The head information comes from the light verb, while the subcat information comes from the verbal noun. The verbal noun has a sub-syntactic status, realized by the [BAR–] attribute.

*Example 19.*

*ryoukai shimashita*  
agreement do-past  
(okay.<sup>18</sup>)

*Example 20.*

*pikkuappu shita*  
pick up do-past  
(picked up<sup>19</sup>)

*Example 21.*

*onagai shimasu*  
beg do-present  
(please<sup>20</sup>)

## 5 Honorification

Spoken language encodes references to the social relation of the dialog partners. The utterances can express social distance between addressee and speaker and third persons, who are mentioned. Honorifics can even express respect concerning entities

<sup>18</sup> From Verbmobil CD27 dialog j046ac.trl.

<sup>19</sup> From Verbmobil CD27 dialog j029ac.trl.

<sup>20</sup> From Verbmobil CD25 dialog j003ac.trl.

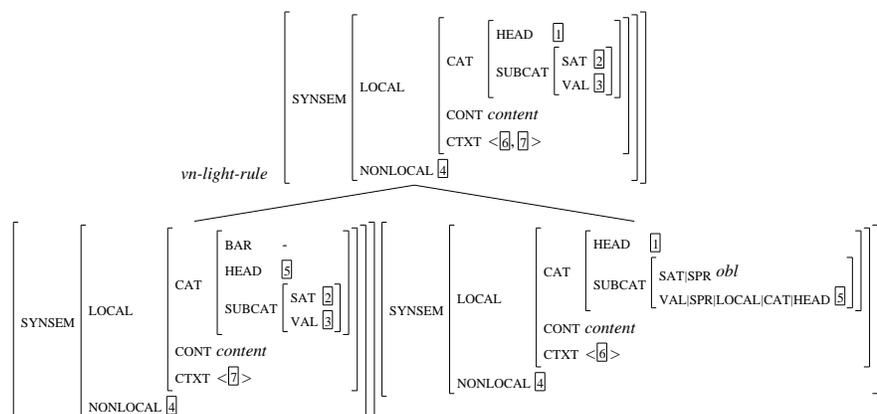


Fig. 7. vn-light rule

of the world. Consider the Example 22. The semantic content of this utterance is: 'When does it suit you?'. But there is an additional pragmatic content: The speaker expresses social distance concerning the addressee. This is expressed by the following attributes:

- The honorific prefix *go* in front of *tsugoo*
- The honorific adjective *yoroshii*
- The honorific copula *deshoo*

Example 22.

*itsu gotsugou ga yoroshii deshoo ka*  
 when HON-conditions NOM good COP QUE  
 (When does it suit you?<sup>21</sup>)

Information about honorification is—on the one hand—necessary for the description of syntactic phenomena like honorific agreement or relative sentences and—on the other hand—necessary for correct translation. In order to understand the whole meaning of the Japanese utterances it is important to represent the different honorific attributes in the analysis structure. The information can be used to resolve zero pronominalization and topicalized structures. It is even more important for the adequate *generation* of the Japanese utterances. In other investigations on zero pronoun resolution in task-oriented dialogs (Siegel, 1996) we calculated that 23.9% of the zero pronouns can be solved using lexical pragmatic restrictions about honorification.

Familiarity or distance between speaker and addressee can be expressed by verbal endings and/or the lexical choice of self-referring pronouns. I will call the relationship of honorifics concerning the relation between speaker and addressee

<sup>21</sup> From Verbmobil CD16, dialog 072.RMN.

**AHON** and give it a polarity  $[-AHON]$  for the plain form in a family context and  $[+AHON]$  for the expressions in a context of social distance.

The social relation between speaker and subject is expressed by the lexical choice of verbs, by the expression *o-VERB-ni-naru*, by the honorific prefix *o/go* at nouns referring to entities belonging to the subject and by the lexical choice of pronouns. I will call this relation between speaker and subject **SHON**. A relation of distance between speaker and subject (where the subject is the addressee or a third person) can be—for example—expressed by the verb *irassharu* (to go), while in a familiar situation the verb *iku* with the same semantic content is used. This is expressed by  $[+SHON]$  and  $[-SHON]$ , respectively. Possible referring expressions for the second and third person can be, for example, *sochira* and *X-san* in relations of distance and *kimi* or *X-kun* in relations of familiarity.

The third relation is the one between speaker and objects in the sentence (other than subject). I will call this relation **OHON**. It is expressed by the lexical choice of these entities and by the honorific prefixes *o* and *go*.

In many cases utterances contain multiple honorification as can be seen in the Example 23. The verbal stem *itashi* expresses subject honorification, the verbal ending *mashi* and the pronoun *watakushi* addressee honorification.

Example 23.

*watakushi ga o-denwa itashi-mashi-ta*  
 I NOM telephone do(hon)-hon-Past  
 (I made a telephone call.)

Japanese honorification undergoes different kinds of restrictions. The first kind to mention is called “pragmatic agreement” by Pollard and Sag (1994). There must be agreement between the SHON honorification of the subject and the verb. Another kind of restriction concerns relative sentences as opposed to complement sentences. Complement sentences allow a honorific predicate (addressee honorification), while relative sentences do not.

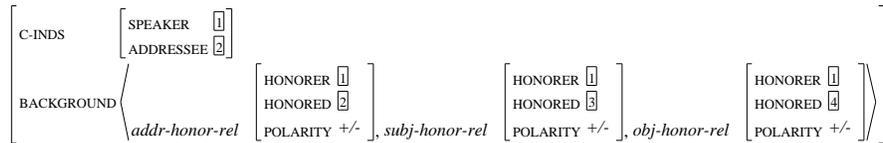


Fig. 8. BACKGROUND in the Verbmobil grammar of Japanese

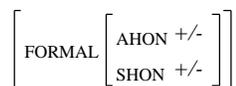
To account for the fact that Japanese honorification has three dimensions, we propose the CONTEXT feature structure in Figure 8. The C-INDS contain indices for speaker and addressee, as proposed by Pollard and Sag (1994). The value of BACKGR is a difference list that sums up the occurring honorificational relations in the utterance. Each occurring relation gets classified into *addr-honor-rel*, *subj-honor-rel* and *obj-honor-rel*. The HONORER is co-indexed with the speaker in all

cases here. The HONORED value is co-indexed with the addressee in C-INDS in the *addr-honor-rel* case, with the subject's CONTENT.INDEX value in the *subj-honor-rel* case and with the CONTENT.INDEX value of the argument that introduces the relation in the *obj-honor-rel* case. The relations all get a value of POLARITY, to account for the fact that there can be forms that are honorific, humble or neutral. A negative SHON polarity, e.g., reflects the situation where the speaker or a third person that socially belongs to the inner circle of the speaker is the subject of the utterance. McGloin (1987) describes this situation socio-linguistically as “positive politeness”, because it expresses social closeness.

Gunji (1987) gives reasons for a syntactic approach<sup>22</sup>. He describes in his JPSG account of Japanese syntax honorification as a kind of agreement:

Since Japanese does not have syntactic agreement phenomena such as number, person, etc., the honorification system is more or less a counterpart.

We expanded the syntactic part of the representation of honorification. The lexical entries get a HEAD feature called FORMAL (see Figure 9). Only the connection of representing honorification on the syntactic **and** contextual level makes it possible to account for all phenomena. The pure syntactic representation cannot account for the representation of honorificational relations between speaker and addressee, for OHON and for multiple honorifications, while the pure contextual representation cannot account for the syntactic restrictions on subjects and relative sentences. The CONTEXT level gives information about felicity of an utterance, while the CAT level gives information about syntactical correctness of an utterance. For honorification in Japanese, we need both. With the fundamental concept of HPSG, the sign, it is possible to incorporate **both** levels of linguistic analysis.



**Fig. 9.** The Head feature FORMAL

Being a HEAD feature, the value of FORMAL is passed up from head daughters to mother daughters. A honorific noun therefore contains the value *SYNSEM|LOCAL|CAT|HEAD|FORMAL|SHONplus*, as well as a verb with subject honorification. For Japanese, we set up the *principle of subject honorification*:

In a honorific lexical structure, the *FORMAL|SHON* value of the HEAD is identical to the *FORMAL|SHON* value of the subject's HEAD and the polarity of the *subj-honor-rel* in BACKGR. The values of the subject's *CONTENT|INDEX* and the HONORED of the *subj-honor-rel* in BACKGR are identical.

<sup>22</sup> See also Ikeya (1983).

An effect of the approach for the machine translation system is that lexical pragmatic restrictions for zero pronouns can be directly accounted for in the analysis. They are essential to find referents for many zero pronouns, as is shown by Metzger and Siegel (1994). Two occurrences of subject honorification can also be possible: Imagine a sentence where the complement sentence has a different subject honorification from the matrix sentence, e.g. in a sentence with indirect speech. Syntactic restrictions for relative sentences can easily be formulated in a way that only verbs with a non-addressee-honorific form can modify nouns.<sup>23</sup>

## 6 Segments and Phrases

As the grammar was developed for the purpose of machine translation of spoken dialogs, it is capable of dealing with spoken language phenomena and ungrammatical or corrupted input. This leads to the necessity of robustness. Being used in an MT system for spoken language, the grammar must accept fragmentary input and be able to deliver partial analyses, where no spanning analysis is available. A complete fragmentary utterance could, e.g., be:

*Example 24.*

*iNtaashitiihoteru*  
Intercityhotel

This is just a noun. There is still an analysis available. If another utterance is corrupted by not being fully recognized, the grammar delivers analyses for those parts that could be understood. An example would be the following best hypothesis from the speech recognizer in a system test:

*Example 25.*

*sou desu ne watakushi no hou wa daijoubu desu da ga kono*  
so COP TAG I GEN side TOP okay COP but this  
*hi wa kayoubi desu ne*  
day TOP Tuesday COP TAG  
(lit.: Well, it is okay for my side, but this day is Tuesday, isn't it?)

Here, analyses for the following fragments are delivered (where the parser found *opera wa* in the word lattice, but not in the hypothesis):

*Example 26.*

- a) *sou desu ne watakushi no hou wa daijoubu desu*  
so COP TAG I GEN side TOP okay COP  
(Well, it is okay for my side.)
- b) *opera wa*  
opera TOP

<sup>23</sup> See also Siegel (2000) for the treatment of Japanese honorification.

- (The opera)
- c) *kono hi wa kayoubi desu ne*  
 this day TOP Tuesday COP TAG  
 (This day is Tuesday, isn't it?)

Another necessity for partial analysis comes from real-time restrictions imposed by the Verbmobil system. If the parser is not allowed to produce a spanning analysis, it delivers best partial fragments.<sup>24</sup>

## 7 Evaluations

The Japanese HPSG grammar in Verbmobil consists of 38 rule schemata (24 unary, 14 binary), 1,184 types and a lexicon of 3399 entries<sup>25</sup>.

We executed a first parsing test on 2607 utterances out of 100 dialogs in the Verbmobil scenario of appointment scheduling. The average sentence length was 8.99 words. Parsing was interrupted, when the chart contained more than 20000 items. We processed only the first reading of every utterance. The result can be seen in Table 1. 2044 utterances (78.4%) got a parsing result, i.e. a spanning analysis. On average, there were 100.60 lexical items processed for every utterance.

**Table 1.** Test of 2607 utterances in Verbmobil scenario I

Phenomenon	total items #	positive items #	word string $\phi$	lexical items $\phi$	parser analyses $\phi$	total results #	overall coverage %
<b>Total</b>	<b>2607</b>	<b>2607</b>	<b>8.99</b>	<b>100.60</b>	<b>1.00</b>	<b>2044</b>	<b>78.4</b>

Next, a parsing test on the utterances of 50 dialogs in the Verbmobil scenario of travel planning was executed. The test contained 7969 utterances of spontaneous language. This test processed exhaustive parsing. The average length of the utterances was 6.22 words. There were on average 76.14 lexical items per utterance. The results can be seen in Table 2. 5807 (72.9%) got a parsing result, i.e. a *spanning analysis*. On average, every utterance got 25.65 parser analyses.

Overgeneration was tested with 223 ungrammatical sentences. Table 3 shows that 7.2% of these got a parsing result.

Table 4 shows the performance of the parser in these tests. They were executed on a 44 MHz Ultra Sparc 2 with Solaris 2.6.

A Verbmobil system test with 491 sentences (all words known, string input) showed no cases, where the Japanese syntax delivered nothing. This shows that in almost any case the syntax module delivers at least partial analyses.

<sup>24</sup> see Kiefer et al., 2000 for further details.

<sup>25</sup> The Japanese HPSG lexicon contains fullforms. The number of lexical entries reflects the high rate of ambiguity of Japanese lexical items, as there are only 2566 word forms in the Japanese word list.

**Table 2.** Test of 7969 utterances in Verbmobil scenario II

Phenomenon	total items	positive items	word string	lexical items	parser analyses	total results	overall coverage
	#	#	$\phi$	$\phi$	$\phi$	#	%
<b>Total</b>	<b>7969</b>	<b>7969</b>	<b>6.22</b>	<b>76.14</b>	<b>25.65</b>	<b>5807</b>	<b>72.9</b>

**Table 3.** Test for overgeneration

Phenomenon	negative items	word string	lexical items	parser analyses	total results	overall coverage
	#	$\phi$	$\phi$	$\phi$	#	%
<b>Total</b>	<b>223</b>	<b>7.40</b>	<b>80.64</b>	<b>11.50</b>	<b>16</b>	<b>7.2</b>

**Table 4.** Performance

	Test 1 (nonexhaustive)	Test 2 (exhaustive)	Test 3 (ungrammatical)
<b>average cpu time</b>	<b>1.74s</b>	<b>2.13s</b>	<b>1.26s</b>

## 8 Conclusion

We described the Verbmobil HPSG grammar for Japanese. This is a robust and efficient grammar for Japanese spoken language. It deals with basic phenomena of Japanese like various nominal and verbal constructions as well as phenomena of spoken Japanese like topicalization, honorification and zero pronouns. Being part of a machine translation system for dialogs, it contains a clear concept of segments and phrases.

## References

- Dubinsky, S. (1997). Syntactic Underspecification and Light-Verb Phenomena in Japanese. *Linguistics* 35(4):627–672.
- Grimshaw, J., and Mester, A. (1988). Light Verbs and Theta-Marking. *Linguistic Inquiry* 19(2):205–232.
- Gunji, T. (1987). *Japanese Phrase Structure Grammar*. Dordrecht: Reidel.
- Ikeya, A. (1983). Japanese Honorific Systems. In *Seoul Papers in Formal Grammar Theory. Proceedings of the 3rd Korean-Japanese Joint Workshop*. Seoul: Hanshin Publishing Company.
- Kiefer, B., Krieger, H.-U., and Siegel, M. (2000). An HPSG-to-CFG Approximation of Japanese. In *Proceedings of Coling 2000*.
- McGloin, N. H. (1987). The Role of WA in Negation. In Hinds, J., Maynard, S.K., and Iwasaki, S., eds., *Perspectives on Topicalization. The Case of Japanese 'WA'*. Amsterdam: John Benjamins Publishing Company. 165–184.
- Metzing, D., and Siegel, M. (1994). Zero Pronoun Processing: Some Requirements for a Verbmobil System. Verbmobil-Memo 46, Universität Bielefeld.

- Nightingale, S. (1996). An HPSG Account of the Japanese Copula and Related Phenomena. Master's thesis, University of Edinburgh.
- Pollard, C., and Sag, I. (1994). *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Siegel, M. (1996). *Die Maschinelle Übersetzung aufgabenorientierter japanisch-deutscher Dialoge. Lösungen für Translation Mismatches*. Ph.D. Dissertation, Universität Bielefeld.
- Siegel, M. (1998). Japanese Particles in an HPSG Grammar. Verbmobil-Report 220, Universität des Saarlandes.
- Siegel, M. (1999). The Syntactic Processing of Particles in Japanese Spoken Language. In Wang, J.-F., and Wu, C.-H., eds., *Proceedings of the 13th Pacific Asia Conference on Language, Information and Computation*, 313–320.
- Siegel, M. (2000). Japanese Honorification in an HPSG Framework. In Ikeya, A., and Kawamori, M., eds., *Proceedings of the 14th Pacific Asia Conference on Language, Information and Computation*, 289–300. Waseda University International Conference Center, Tokyo: Logico-Linguistic Society of Japan.