

Head-Initial Constructions in Japanese

Melanie Siegel
University of
Saarbrücken
Germany

siegel@dfki.de

Emily M. Bender
University of
Washington
USA

ebender@u.washington

1 Introduction

Japanese is generally taken to be strictly head-final in its syntax (Gunji 1987). In our work on a broad-coverage, precision implemented HPSG for Japanese, we have found that while this is generally true, there are nonetheless a few minor exceptions to the broad trend. In this paper, we describe the grammar engineering project (Section 2), present the exceptions we have found (Section 3), and conclude that this kind of phenomenon motivates on the one hand the HPSG type hierarchical approach which allows for the statement of both broad generalizations and exceptions to those generalizations and on the other hand the usefulness of grammar engineering as a means of testing linguistic hypotheses (Section 4).

2 The JACY grammar

Our Japanese HPSG grammar originates from work done in several research projects concerning different domains. The grammar is couched in the theoretical framework of Head-Driven Phrase Structure Grammar (HPSG) (Pollard & Sag 1994), with semantic representations in Minimal Recursion Semantics (MRS) (Copestake et al. 2001). We also use the ChaSen tokenizer and POS tagger (Matsumoto et al. 2000).

All research projects the grammar was embedded in share the following characteristics: The grammar was deployed in practical applications, and developed to handle large and realistic corpora. The domain focus is spoken or close-to-spoken language. This requires the treatment of core as well as peripheral phenomena of the language. In extending coverage to more peripheral phenomena, we have found some which are best treated as head-initial, including both head-complement constructions (number names and certain uses of numeral classifiers) and head-modifier constructions (head-initial

modification of nouns, postpositional phrases, verbs and temporal expressions).

The grammar implementation is based on a system of types. There are 900 lexical types that define the syntactic, semantic and pragmatic properties of the Japanese words, and 188 types that define the properties of phrases and lexical rules. The grammar includes 50 lexical rules for inflectional and derivational morphology and 47 phrase structure rules. The lexicon contains about 30000 stem entries and a mechanism to assume default lexical types for items that can be POS tagged by ChaSen, but are not included in the HPSG lexicon.

JACY is open-source and downloadable from <http://www.dfki.de/~siegel/grammar-download/JACY-grammar.html>.

3 The position of syntactic heads in Japanese

The syntactic head of a construction is that element which determines the syntactic distribution of the whole.¹ This notion of head is, of course, fundamental to HPSG, as encoded in the head-feature (Pollard/Sag 1994) and subcategorization (Borsley 1993) principles. The HEAD values encode precisely the kind of part of speech information which determines the syntactic distribution of an element (such as case, preposition form, and modification possibilities), and the head feature principle propagates this information to the mother of the phrase. Likewise, the subcategorization principle distinguishes heads from arguments.

Deciding on the head constituent in a phrase therefore means to observe which constituent contributes the head information and the subcategorization information.²

Zwicky (1993) gives the following table to differentiate heads and dependents.³

¹ Note that the syntactic head need not be the semantic head.

² Although argument transfer and composition in constructions like verbal noun – light verb underly different restrictions for subcategorization.

³ In modifier constructions, the semantic functor is not the head, but the modifier. Cf Zwicky 1993.

	<i>Head</i>	<i>Dependent</i>
Semantics	characterizing	contributory
Syntax	required	accessory
	Word rank	Phrase rank
	category determinant	non-determinant
	external representative	externally transparent
Morphology	morphosyntactic locus	morphosyntactically irrelevant

By this definition, it is true that most heads in Japanese follow both arguments and adjuncts: Verbs appear at the end of clauses, adjectives, genitives, and relative clauses precede nouns, and the language has postpositions, including both 'contentful' elements such as *kara* 'from', and the case marking postpositions *ga*, *wo*, *ni*. The contrast in example 1 illustrates how the distribution of a phrase ending in *-ga* is determined by *-ga*:

example 1

a) nanji kara ga yoroshii desu ka?
when from CASE good COP interr.

(From what time on would be good?)

b) nanji kara desu ka?
when from COP interr.

(From when is it?)

c) *nanji kara ga desu ka?
**when from CASE COP interr.*

We now turn to the exceptions we have found to the general head-final trend, which can be classified into two groups:

3.1 Head-initial modification

Using the definition above of the syntactic head in a construction, we can find some elements that behave as non-heads, although they occur final in a construction. The first group is certain elements which modify nouns. The modifier *dake* 'only' occurs between nouns and case particles, as for example in:

example 2

Nomura-san dake ga kita
Ms. Nomura only CASE came
 (Only Ms. Nomura came)

The head of the construction *Nomura-san dake ga* is the case particle *ga*, because the verb *kita* selects for a subject marked by *ga* and therefore *ga* contributes the information for syntactic selection. The head of *Nomura-san dake* must be *Nomura-san*, because *ga* selects for a noun. Leaving *dake* out in this construction leads to a grammatical sentence *Nomura-san ga kita*, while leaving *Nomura-san* or the case particle out, leads to ungrammatical sentences.⁴ Therefore we conclude that *dake* in this construction is a modifier to *Nomura-san* and we have a good example of head-initial noun modification. A second element, *nomi* 'only', is very similar to *dake*, except that it cannot follow adjectives and quantifiers. It is used in formal speech and written Japanese, but more seldom in the registers found in our corpora. Temporal expressions also take certain post-head modifiers. Here is an example from Mainichi Shinbun:

example 3

Tokyo kara kuruma de nijikan
Tokyo from car with 2 hours

bakari no kinkou no onsen ni
about Gen suburb Gen hot spring to

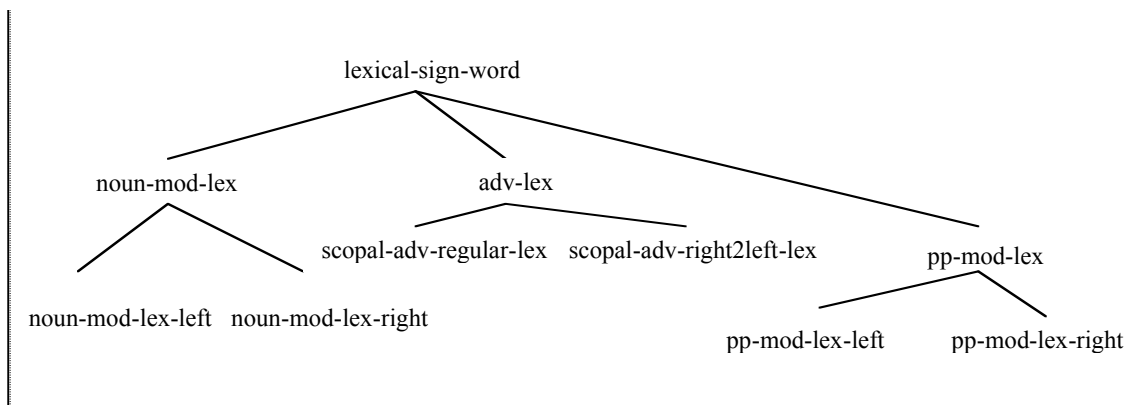
asa shichiji goro shuppatsu-suru
morning 7o'clock about depart

(We drive to the hot spring in the suburbs, which is about two hours away from Tokyo, at about 7 a.m.)

The relevant construction here is *nijikan bakari no*. The head of the construction is *no*, because it carries the information that the construction can modify an NP. *no* on the other hand, selects for the noun *nijikan* that is modified by *bakari*. The sentence would be perfectly grammatical without *bakari*.

Also for *goro*, *kurai* and *hodo* (about), one finds several examples for head-initial modification of temporal expressions, such as:

⁴ Although particle omissions can occur in spoken language.



example 4

kyou nanji goro made
today what time about until
 nete-imashita ka
slept question

(Until about what time did you sleep today?)

We also find post-head modifiers of PPs such as *bakari* ‘only’ and *dake* ‘only’.

example 5

shoutotsu ni bakari kanshin ga
collision to only concern CASE

atsumatta
collected

(The concern is concentrated only to collision.)

example 6

riyousha wa Tokyo kara dake
user TOPIC Tokyo from only
 de-wa-nai
not to be

(The users were not only from Tokyo)

In these examples, the particles *ni* ‘to’ and *kara* ‘from’ determine the combinatorical potential of the whole phrase, leaving *bakari* and *dake* the role of modifiers.

There are also examples of head-initial verb modification. Here is one from Mainichi Shinbun (2002):

example 7

gakkou no Sensei wo okorasete
schook GEN Teacher CASE upset
 bakari ita
only AUX

(The only thing he was doing was upsetting the teacher)

This is one exception to the general rule that nothing should interfere between a verb in *te* form and an auxiliary. The exception can be explained, if *bakari* modifies *sabotte*. We therefore introduce one instance of *bakari* (similarly, one for *dake*) that can be a post-head modifier of verbs with *te* inflection.

Our analysis for head-initial modification consists of:

1. A lexical type hierarchy containing types that allow for head-initial constructions.
2. Grammar rules for head-initial modification and head-initial complementation.
3. A head feature POSTHEAD that is referenced by head-adjunct rules.

Figure 1 shows part of the type hierarchy of lexical signs, containing lexical items that modify nouns, postpositions and verbs, and which are divided into left-modifying and right-modifying items.

The inventory of grammar rules contains rules for both head-initial and head-final complementation, which differ in the order of the daughters. The rules reference the HEAD.POSTHEAD value of the modifier daughter in order to constrain the distribution of lexical items across the constructions. POSTHEAD can be *left* or *right*, or can be left

unspecified for those items that can modify in both directions.⁵

3.2 Head-initial complementation

We have found two clear cases of head-initial complementation, the first in number names and the second in numeral classifiers. In both cases, one optional argument follows the head.

We argue that number names like *ni hyaku juu* '210' are head-medial on the basis of examples like 8 and 9. The 8b and 8c each share one element in common with 8a. The examples in 9 show that the external distribution of these phrases differ.

example 8

- a) ni hyaku juu
two hundred ten
- b) go hyaku san
five hundred three
- c) ni sen san
two thousand three

example 9

- a) roku sen ni Hyaku juu
six thousand two Hundred ten
- b) roku sen go Hyaku San
six thousand five Hundred three
- c) *roku sen ni sen san
*six thousand two thousand three
- d) *roku sen Go sen Juu
*six thousand Five thousand ten

Expressions with *hyaku* (8a and 8b) have the same combinatoric potential. Expressions without *hyaku* differ. The other elements of example 8 *ni* 'two' and *juu* 'ten' are not

relevant. Thus, we take *hyaku* to be the head of example 8. If we forget for the moment that Japanese is supposed to be head-final, this isn't very surprising. English number names work the same way (see Smith 1999). So do number names in another SVO language: Chinese, the source from which Japanese borrowed this system.

One might argue that this is actually a morphological process, in which case the head-medial structure is less surprising. However, Martin (1987) finds that some local combinations within number names (e.g., the names for 11 through 19, 20, 30, 200, 300, etc.) form single phonological words, while longer combinations made up of these pieces (such as *sanbyaku juuichi* '311') show phrasal phonology. The analysis presented here was developed within the context of an application that takes text-based input. As such, it was most convenient to apply the phrasal analysis uniformly. A similar analysis could be developed that provides lexical entries for every combination that forms a phonological word.

The second type of head-initial complementation involves numeral classifiers. All numeral classifiers combine with a number name to their left, but certain mensural numeral classifiers such as *nen* 'year' can also take the word *han* 'half' to their right. (see 10) Syntactically, the numeral classifier determines the combinatorics of the phrase (being able to modify nouns, not being able to show up as the specifier of a larger number name). The presence or absence of *han* has no effect on the distribution.

example 10

- a) ni nen han
two years half
- b) ni nen
two years

Our analysis of both of these instances of head-initial complementation consists of:

- 1) two head-complement rules, differing in the order of the daughters, and

⁵ We also use POSTHEAD for the selection of relative sentence constructions, coordinated structures and the head selection of nominal compounds (see Radford 1993 for criteria on head selection in nominal compounds).

- 2) a high-level distinction in the subtypes of *head* into *init-head* and *final-head*

The two head-complement rules are sensitive to the head type of their head daughter. Most head types are subtypes of *final-head*, giving the general pattern, while numeral classifiers and number names are given subtypes of *init-head*.

4 Conclusion

We believe that the rather peripheral exceptions noted here do not detract from the broad generalization that Japanese has a very strong tendency to be head-final. Rather, they illustrate once again the fact that languages seamlessly combine general tendencies with particular exceptions. In order to build consistent grammars that scale up to ever larger fragments of the languages we wish to model (such as is required for practical applications), we require a framework that allows the statement of generalizations at varying degrees of granularity. Furthermore, we believe that the construction of broad-coverage precision grammars such as JACY in the context of applications which require robustness in the face of real-world language use provides a useful discovery procedure for many of the smaller generalizations and exceptional cases (cf Baldwin et al 2004).

5 References

- Baldwin, T., J. Beavers, E.M. Bender, D. Flickinger, A. Kim, & S. Oepen (2004):** *'Beauty and the Beast: What running a broad-coverage precision grammar over the BNC taught us about the grammar – and the corpus.'* Paper presented at the International Conference on Linguistic Evidence, Tübingen, Germany.
- Borsley, R. D. (1993):** Heads in HPSG. In Greville Corbett and N. Fraser and S. McGlashan, editor(s), *Heads in Grammatical Theory*.
- Copetake, A., Lascarides, A. and D. Flickinger (2001):** An Algebra for Semantic Construction in Constraint-based Grammars. In: *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL 2001)*, Toulouse, France.
- Radford, A. 1993):** Head-hunting: on the trail of the nominal Janus. In Greville Corbett and N. Fraser and S. McGlashan, editor(s), *Heads in Grammatical Theory*.
- Gunji, T. (1987):** *Japanese Phrase Structure Grammar*. Dordrecht: Reidel.
- Martin, S.E. (1987):** A reference grammar of Japanese. Tokyo.
- Matsumoto, Y., Kitauchi, A., Yamashita, T., Hirano, Y., Matsuda, H., Takaoka, K., Asahara, M. (2000):** *Morphological Analysis System ChaSen version 2.2.1 Manual*.
- Pollard, C. & I.A. Sag (1994):** *Head-Driven Phrase Structure Grammar*. Chicago/London.
- Smith, J. D. (1999):** English number names in HPSG. In: G. Webelhuth, J-P. Koenig, and A. Kathol (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, 145-160. Stanford, CA: CSLI.
- Zwicky, A.M. (1993):** Heads, bases and functors. In Greville Corbett and N. Fraser and S. McGlashan, editor(s), *Heads in Grammatical Theory*.