Head-Initial Constructions in Japanese

Melanie Siegel
University of Saarbrücken

Emily M. Bender
University of Washington

Proceedings of the 11th International Conference on
Head-Driven Phrase Structure Grammar

Center for Computational Linguistics, Katholieke Universiteit Leuven

Stefan Müller (Editor)

2004

CSLI Publications

pages 244–260

http://csli-publications.stanford.edu/HPSG/2004

Siegel, Melanie, & Bender, Emily M. 2004. Head-Initial Constructions in Japanese. In Müller, Stefan (Ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, 244–260. Stanford, CA: CSLI Publications.

Abstract

Japanese is often taken to be strictly head-final in its syntax. 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, present the exceptions we have found, 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.

1 Introduction

Japanese is generally taken to be strictly head-final in its syntax (Gunji, 1987). Broad claims like this can be tested by implementing grammars for large fragments of the language and testing them against naturally occurring text. In our work on a broad-coverage, precision implemented HPSG for Japanese, we have found a few minor exceptions to the broad trend towards head-final order in Japanese. The remainder of this section describes the grammar engineering project and the kinds of linguistic data so far considered. Section 2 discusses how to identify heads and presents the general trend towards right-headedness in Japanese syntax. Sections 3 and 4 describe and motivate the two kinds of exceptions we have found: head-initial modification and head-initial complementation. In Section 5 we 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. The analyses are presented in greater detail in the appendix.

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 and Sag, 1994), with semantic representations in Minimal Recursion Semantics (MRS) (Copestake et al., 2003). We use the ChaSen tokenizer and POS tagger (Matsumoto et al., 2000). In the context of a broader multilingual grammar engineering effort, JACY has been made compatible with the Grammar Matrix (Bender et al., 2002), including incorporating Matrix types. One of the benefits of Matrix-compatibility is that the MRS representations produced by the grammar are consistent with those produce by Matrix-derived grammars for other languages. This improves interoperability with back-end systems.

[†]We would like to thank Atsuko Shimada for assistance with data questions and the JACY grammar in general; Francis Bond for his work on JACY; YY Technologies, DFKI, and Project DeepThought for opportunities to develop the grammar; and three anonymous reviewers for the HPSG conference for valuable suggestions and comments. All remaining errors are our own.

JACY was first implemented for the use in Machine Translation of spoken dialogs (Verbmobil; Siegel, 2000). The next application area was interpreting email for automated response (Oepen et al., 2002). Project DeepThought[1] embedded the Japanese grammar in a multilingual grammar development framework for hybrid natural language processing (Uszkoreit et al., 2004). All of these development contexts share the following characteristics: (i) The grammar was deployed in practical applications and developed to handle large and realistic corpora. (ii) The domain focus is spoken or near-spoken language. Such development contexts require 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 31 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/JACYgrammar.html.

2 The position of syntactic heads in Japanese

Zwicky (1993) identifies several characteristics which have been taken to differentiate heads and dependents, and points out that they do not correlate all that well.[2]

	Head	Dependent
Semantics	characterizing	contributory
Syntax	required	accessory
	word rank	phrase rank
	category determinant	non-determinant
	external representative	externally transparent
Morphology	morpho-syntactic locus	morpho-syntactically irrelevant

Table 1: Characteristics of heads and dependents, from Zwicky 1993

HPSG theory only recognizes some of these characteristics in the identifica-

[1] http://www.project-deepthought.net

[2] In modifier constructions, the semantic functor is not the head, but the modifier, cf. Zwicky 1993.

tion of syntactic heads,[3] namely required v. accessory, category determinant v. non-determinant, and external representative v. externally transparent. The central intuition is that the syntactic head of a construction is that subconstituent which determines the syntactic distribution of the whole.

This notion of head is, of course, fundamental to HPSG and is encoded in the head-feature (Pollard and Sag, 1994) and subcategorization (Borsley, 1993) principles. Given an HPSG grammar, the head of any constituent parsed by the grammar is well-defined. 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, in general making the valence requirements of a phrase some function of the valence requirements of its head.[4] Determining which element is the head for the purposes of writing the grammar, on the other hand, can be trickier. Deciding on the head constituent in a phrase requires observing which constituent contributes the head information and the subcategorization information.

By this definition, it is true that most heads in Japanese follow both arguments and adjuncts: Verbs appear at the end of clauses, as can be seen in example (1).[5]

(1) Tanaka ga hon wo yonda.
Tanaka NOM book ACC read.past
'Tanaka read a book.'

Adjectives, genitives, and relative clauses precede nouns:

(2) Tanaka no yasashii tomodachi ga kita.
Tanaka GEN nice friend NOM come.past
'Tanaka's nice friend came.'

The language has postpositions, including both contentful elements such as *kara* 'from' (3), and the case marking postpositions *ga, wo, ni* (4), which both follow nouns.

(3) Toukyou kara kita
Tokyo from come.past
'(someone) came from Tokyo.'

[3]Note that the syntactic head need not be the semantic head.

[4]In some cases these 'functions' get fairly elaborate and also refer to the valence requirements of the non-head daughter, as in argument transfer and composition in constructions like that combining verbal nouns and light verbs in Japanese.

[5]Abbreviations used in this paper are as follows: ACC accusative, AUX auxiliary, COP copula, GEN genitive, INSTR instrumental, LOC locative, NEG negative, NOM nominative, NUMCL numeral classifier, TOP topic, Q interrogative particle.

(4) Nanji kara ga yoroshii desu ka?
What time from NOM good COP Q
'From what time would be good?'

That contentful postpositions should head their phrases is relatively uncontroversial. Applying the same treatment to the case markers might be more surprising, especially as they are sometimes considered to be nominal inflection (e.g., Sag et al., 2003). However, Siegel (1999) makes the case for treating them as heads. We illustrate the argument here with the examples in (3)–(5), which show that *ga* is crucial in determining the combinatoric potential of its phrase.

(5) a. Nanji kara atsumarimasu ka?
What time from gather Q
'From what time are people gathering?'

b.*Nanji kara ga atsumarimasu ka?
What time from NOM gather Q

In (4), there is a single constituent (*Nanji kara ga*) containing both a contentful postposition (*kara* 'from') and a case-marking postposition *ga*. Constituents ending in *kara* are verbal adjuncts ((3) and (5a)). When *ga* attaches, the result is eligible to appear in an argument (here, subject) position (4), and no longer can appear as a verbal adjunct (5b). If *ga* were merely a marker that otherwise preserved the category information of the constituent it attaches to, this behavior would be hard to explain. Note that on this analysis, the Japanese case particles look fairly similar to English 'case-marking prepositions', such as *to* in *Kim gave the book to Sandy.*For our purposes here, the main point is that PPs, with both contentful and case marking postpositions, are also head-final.[6]

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

3 Head-initial modification

3.1 Data

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. In this class, we find the modifiers *dake*, *nomi*, *bakari* (in two distinct uses), *goro*, *kurai*, *hodo*, and certain instances of numeral classifiers.

[6]In general, distinguishing morphology and syntax is not very clear-cut in this agglutinating language (Shibatani and Kageyama, 1988; Kageyama, 2001). For better or for worse, the orthography does not provide any clues, lacking inter-word spaces. For practical (engineering) purposes, we tend towards regarding syntax over morphology, as ChaSen provides near-morpheme-level segmentation. Along the way, we will point out evidence that the cases presented here involve syntactically separate words (clitics or otherwise).

3.1.1 *Dake*

The modifier *dake* 'only' modifies at least NPs, predicative PPs, and adverbs.

The noun-modification use is illustrated in (6):

(6) a. Nomura-san dake ga kita.
Ms. Nomura only NOM come.past
'Only Ms. Nomura came'

b. Nomura-san ga kita.
Ms. Nomura NOM come.past
'Ms. Nomura came'

c.*Dake (ga) kita.
only NOM come.past

The head of the construction *Nomura-san dake ga* is the case particle *ga* (see above). 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* out gives an ungrammatical sentence. *Dake* is optional in all registers, the noun is obligatory in all, and the case particle is obligatory in some. Therefore we conclude that *dake* in this construction is a modifier to *Nomura-san*, even though it follows the head.

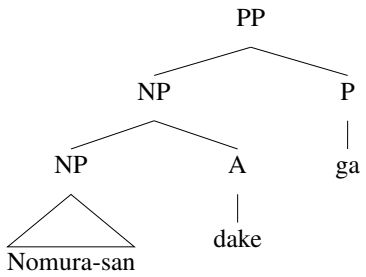

Figure 1: Structure of PP with *dake*

The predicative PPs modifier use of *dake* is illustrated in (7):

(7) Riyousha wa toukyou kara (dake) de-wa-nai.
Users TOP Tokyo from (only) COP-NEG
'The users were not only from Tokyo.'

The fact that dake is optional in this example lends support to the conclusion that *toukyou kara dake* is a head-initial construction. Further support comes from the fact that the order of the particles is flexible, as illustrated in (8) (from Makino and Tsutsui, 1986, 95).

(8) a. Kono kuruma wa arukouru de dake ugokemasu.
This car TOP alcohol INST only move
'This car runs only on alcohol.'

b. Kono kuruma wa arukouru dake de ugokemasu.
This car TOP alcohol only INST move
'This car runs on alcohol alone'

As indicated in the glosses, *dake* can modify (semantically as well as syntactically) either the NP or the PP. It can appear in either position without affecting combinatoric potential. Thus, *arukouru de dake* and *arukouru dake* are head-initial

Finally, adverbs can also be modified (head-initially) by *dake*, as illustrated in (9) (from Makino and Tsutsui, 1986, 94).

(9) Watashi wa nihon e ichido (dake) itta.
I TOP Japan to once (only) went
'I went to Japan (only) once.'

To summarize the observations for *dake*, we can say that it combines with (at least) NP, PP, and ADV to form a category of the same type. The relative non-specificity of the host suggests a syntactic rather than a morphological combination. The distributional facts support treating *dake* as a non-head, even though it is final in its constituent.[7]

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 seldom in the registers found in our corpora.

3.1.2 *Bakari* 'only'

Our second example is *bakari* 'only'. It can modify PPs and VPs (or possibly Vs). Consider first the example in (10a), from the newspaper *Mainichi Shinbun*.[8] Here, *bakari* is a PP modifier:

(10) a. @Shoutotsu ni bakari kanshin ga atsumatta.
collision to only concern *nom* collected
'It is only on the collision that concern is concentrated.'

b. Shoutotsu ni kanshin ga atsumatta.
collision to concern NOM collected
'It is on the collision that concern is concentrated.'

c.*Shoutotsu bakari kanshin ga atsumatta.
collision only concern NOM collected

[7] Makino and Tsutsui (1986) also note a use of *dake* where it attaches to verbs and adjectives to make nominal constituents. In this case, *dake* appears to be a nominalizing head and the examples are not relevant to the point at hand.

[8] Following Bender and Kathol (In press), we mark attested examples with @.

In this example, the particle NI 'to' determines the combinatoric potential of the whole phrase, leaving *bakari* the role of a modifier.

There are also examples of head-initial verb modification, including the following attested in *Mainichi Shinbun* in 2002:

(11) @Gakkou no sensei wo okorasete bakari ita
school GEN teacher ACC upset only AUX
'The only thing he was doing was upsetting the school's teacher.'

This is one exception to the general rule that nothing should intervene between a verb in the *-te* form and an auxiliary. The exception can be handled if *bakari* modifies *okorasete*. We therefore introduce one instance of *bakari* that can be a post-head modifier of verbs with *-te* inflection.[9]

3.1.3 *Bakari* and other forms meaning 'about'

There is another post-head modifier *bakari* meaning 'about', which modifies temporal expressions. We illustrate it here with another *Mainichi Shinbun* example:

(12) @Toukyou kara kuruma de nijikan bakari no kinkou no onsen ni
Tokyo from car INST 2 hours about GEN suburb GEN hotspring to
asa shichiji goro shuppatsu-suru.
morning 7 o'clock around depart
'We depart at about 7 a.m. for a hotspring in the suburbs which is about two hours from Tokyo by car.'

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*, in turn, selects for the temporal noun *nijikan* and *nanjikan* is modified by *bakari*. The sentence would be perfectly grammatical without *bakari*.

Similarly, for *goro, kurai* and *hodo* (about), one finds several examples for head-initial modification of temporal expressions, such as (13):

(13) Kyou nanji goro made nete imashita ka?
today what time about until sleep AUX.past Q
'Until about what time did you sleep today?'

Leaving out *goro* in (14a) simply removes the 'approximate' meaning from the sentence, while leaving out *nanji* (14b) changes the meaning drastically: *Goro* becomes a modifier of *kyou*. Leaving out *made* (14c) gives the sentence another meaning, 'At about what time did you fall asleep today?'. Leaving out both *goro* and *made* gives 'At what time did you fall asleep today?'.

[9] An anonymous reviewer points out that the class of elements which can appear in this position also includes *wa*, *mo*, *dake*, *koso*, *sae*, and *made*. Many, if not all of the elements, should be susceptible to a similar analysis.

(14) a. Kyou nanji made nete imashita ka?
today what time until sleep AUX.past Q
'Until what time did you sleep today?'

b. Kyou goro made nete imashita ka?
today about until sleep AUX.past Q
'Were you sleeping until about today?'

c. Kyou nanji goro nete imashita ka?
today what time about sleep AUX.past Q
'At about what time did you fall asleep today?'

d. Kyou nanji nete imashita ka?
today what time sleep AUX.past Q
'At what time did you fall asleep today?'

Once again, we see a modifier (*goro*) which can attach to multiple different constituents. Unlike *made*, *goro* does not affect the way the constituent it is attached to interacts with the rest of the sentence. Therefore, we propose the structure in Figure 2 for *nanji goro made*.

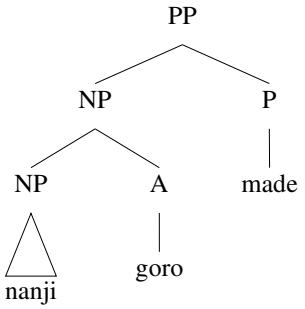

Figure 2: Structure of *nanji goro made* 'until about what time'

3.1.4 Numeral classifiers

Finally, on our analysis, numeral classifier phrases appearing between a noun and its case particle or immediately after a case particle are post-head modifiers. Some examples are given in (15). See Bender and Siegel, 2004 for further details.

(15) a. Neko ni hiki wo kau.
cat 2 NUMCL ACC raise
'(I) am raising two cats.'

b. Neko wo ni hiki ie de kau.
cat ACC 2 NUMCL house LOC raise
'(I) am raising two cats in my house.'

3.1.5 Summary

In this section, we have seen post-head modification of nominal, postpositional, adverbial and verbal constituents. Many of the modifiers can modify multiple different parts of speech. Others (numeral classifier phrases) are internally complex (potentially containing arbitrarily large number names) and further more can appear before or after the phrases they modify, or 'floated' away from them (Bender and Siegel, 2004). These properties suggest that we are dealing with a syntactic rather than morphological phenomenon.

3.2 Analysis

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 3 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.

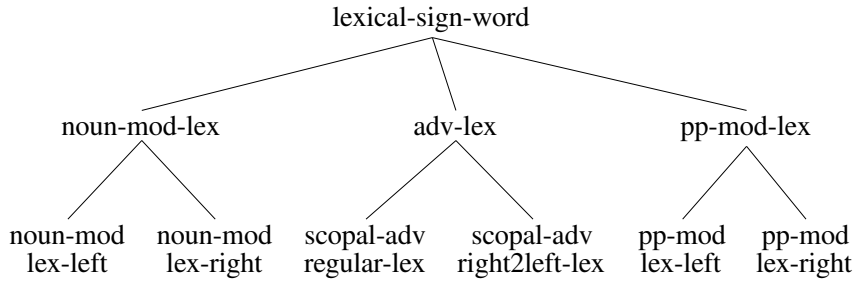

Figure 3: Partial hierarchy of lexical types for modifiers

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. [10] The details of the analysis are presented in the appendix.

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

4 Head-initial complementation

4.1 Data

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 (16) and (17). (16b) and (16c) each share one element in common with (16a). The examples in (17) show that the external distribution of these phrases differ.

(16) a. ni hyaku juu
two hundred ten

b. go hyaku san
five hundred three

c. ni sen san
two thousand three

(17) 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* ((16a) and (16b)) have the same combinatoric potential. Expressions without *hyaku* differ. The other elements of (16) *ni* 'two' and *juu* 'ten' are not relevant. Thus, we take *hyaku* to be the head of (16). 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 while some local combinations within number names (e.g., the names for 11 through 19, 20, 30, 200, 300, etc.) form single phonological words, longer combinations made up of these pieces (such as *sanbyaku juuichi* '311') show phrasal phonology.

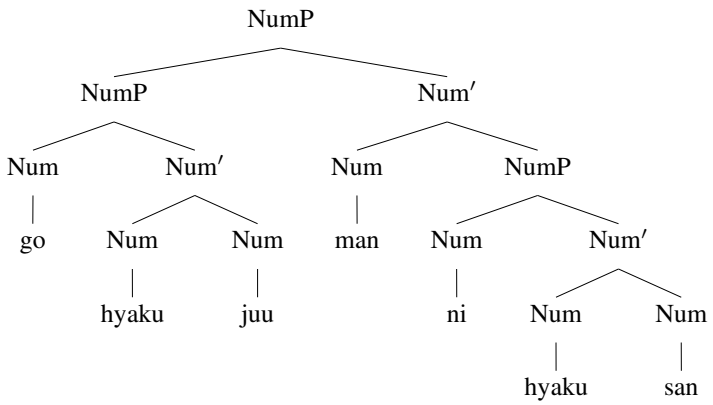

Figure 4: Center recursion in number name expressions

Moreover, number names shows the sort of center recursion that distinguishes context-free languages from regular languages (see Figure 4). This kind of recursion is (to our knowledge) unattested elsewhere in morphology.

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. It would still involve head-initial structures: In a phrase like *sanbyaku juiuichi*, the phonological words are *sanbyaku* ('three hundred') and *juuichi* ('eleven'). Following the same argumentation as above, *sanbyaku* (and within it, *hyaku*, meaning 'hundred') determines the distribution of the phrase within larger number names.

(18) a. issen [sanbyaku juuichi]
one thousand three hundred eleven

b.*gohyaku [sanbyaku juuichi]
five hundred three hundred eleven]

c. gohyaku [juuichi]
five hundred eleven]

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 (19). 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. The numeral classifier is also in a better position to integrate the semantics of *han* than vice versa (Bender and Siegel, 2004).

(19) a. ni nen han
two years half

b. ni nen
two years

4.2 Analysis

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 sensitive to the HEAD type of the head
2. A high-level distinction in the sub-types 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*. The details of this analysis are presented in the appendix.

5 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 (cf. Fillmore et al., 1988). 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)).

Appendix: Rules and types

In this appendix, we present the details of the rules and types used to implement these analyses in the JACY grammar. Complete details can be found by downloading JACY.[11] The basic idea is simple: Separate head-complement and head-adjunct rules for each order[12] which are keyed to particular features on one of the daughters: the adjunct in head-adjunct constructions and the head in head-complement

[11] From http://www.dfki.de/ siegel/grammar-download/JACYgrammar.html.
[12] The formalism we work with does not separate ID and LP rules.

constructions. We believe that there are many possible ways of representing the basic claims in this paper. This particular analysis has been integrated into the broad-coverage grammar and tested against corpus data. Thus, we know that it is consistent with a significant fragment of the language.

Head-Initial Modification

As described above, there are three components to this analysis: a lexical type hierarchy (see Figure 3) which allows for head-initial modification, head-initial (as well as head-final) versions of the relevant modifier rules, and a feature POSTHEAD which the rules reference to constrain the distribution of the modifiers.

Thus, head-initial modifier rules (scopal or intersective) bear these constraints, where the feature ARGS encodes the daughters of the rule and the order in which they appear:

(20) $\begin{bmatrix} \text{HEAD-DTR} & \boxed{1} \\ \text{NON-HEAD-DTR} & \boxed{2}\left[\ldots\ \text{CAT.HEAD.POSTHEAD} \quad \textit{right}\right] \\ \text{ARGS} & \langle\, \boxed{1}, \boxed{2} \,\rangle \end{bmatrix}$

Modifiers of type *pp-mod-lex-right*, etc., are constrained to be [POSTHEAD *right*], and are compatible with head-initial modifier rules. In contrast, *pp-mod-lex-left*, etc., are [POSTHEAD *left*] are thus incompatible with head-initial modifier rules. In principle, modifiers could be underspecified for POSTHEAD, thus appearing on either side. Our lexicon does not currently contain any such modifiers.[13]

Head-Initial Complementation

The analysis of head-initial complementation again involves two rules, and a means of restricting heads to one or the other. In this case, we do not posit an additional feature, but instead take advantage of the type hierarchy and posit a split between initial heads and final heads. Most head types inherit from *final_head*, including *noun-or-case-p_head* (subsuming nouns and the case particles), *verb_head*, and *p_head*, for the contentful post-positions. The two subtypes of *init_head* are *int_head* (for number names) and *num-cl_head* (for numeral classifiers). The latter point is a bit subtle: The only numeral classifiers that take complements at all are those that can appear with *han* (as a complement).[14]

[13] It might appear that numeral classifiers would constitute a case of modifiers attaching either to the left or the right of their heads. However, in pre-head uses of numeral classifiers there is always an intervening *no* (genitive) particle. We treat this particle as a head which selects for a numeral classifier phrase and mediates the modification of the noun by the numeral classifier. For details, see Bender and Siegel, 2004.

[14] We have actually found it convenient to posit one more kind of numeral classifier which takes a complement: namely currency symbols such as '$', which appear to the left of a numerical expression but otherwise function syntactically and semantically like currency words such as *doru* and *en*, which appear to the right of a number name. Most numeral classifiers select their dependent number name

As the classification into *final_head* and *init_head* is only referenced by the head-complement rules, it is simplest to make them all *init_head*.

The following constraints on the two head-complement rules capture the necessary contrast:

(21) a. head-complement-head-final-rule:

$$\begin{bmatrix} \text{HEAD-DTR } \boxed{1}\begin{bmatrix}\text{SYNSEM.LOCAL.CAT.HEAD} & \textit{final\_head}\end{bmatrix} \\ \text{NON-HEAD-DTR } \boxed{2} \\ \text{ARGS } \langle \boxed{2}, \boxed{1} \rangle \end{bmatrix}$$

b. head-complement-head-initial-rule:

$$\begin{bmatrix} \text{HEAD-DTR } \boxed{1}\begin{bmatrix}\text{SYNSEM.LOCAL.CAT.HEAD} & \textit{init\_head}\end{bmatrix} \\ \text{NON-HEAD-DTR } \boxed{2} \\ \text{ARGS } \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix}$$

The ordering constraints relating HEAD-DTR, NON-HEAD-DTR, and ARGS are inherited from a supertype that is also applicable to the head-modifier cases.

In our current implementation, there are no head types which are indeterminate between *init_head* and *final_head*. All head types inherit from exactly one of these. It would of course be possible to cross-classify the ordering dimension with the part of speech dimension, should this be necessary, if some elements of a certain head type preceded their complements and others followed or if all elements of some head type could appear in either order with respect to their complements. Our investigations so far suggest that this is not the case for Japanese. It might be relevant for another language with relatively free order in general, but with some heads showing a more fixed order.

References

Baldwin, Timothy, Beavers, John, Bender, Emily M., Flickinger, Dan, Kim, Ara and Oepen, Stephan. 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: Empirical, Theoretical, and Computational Perspectives, Tübingen, Germany.

Bender, Emily M., Flickinger, Dan and Oepen, Stephan. 2002. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars. In John Carroll,

as a specifier. Currency symbols select their dependent number name as a complement. As the complement appears after the head in this case, it is consistent with making all numeral classifiers be *init_head*. However, this pattern is strictly a phenomenon of the written language: when such strings are read aloud, the current symbol is pronounced as a currency word after the number name. Therefore, the status of this pattern with respect to the points made in this paper is murky at best.

Nelleke Oostdijk and Richard Sutcliffe (eds.), *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pages 8–14, Taipei, Taiwan.

Bender, Emily M. and Kathol, Andreas. In press. Constructional Effects of *Just Because . . . Doesn't Mean . . .* . In *BLS27*.

Bender, Emily M. and Siegel, Melanie. 2004. Implementing the Syntax of Japanese Numeral Classifiers. In *Proceedings of IJCNLP-04*.

Borsley, Robert D. 1993. Heads in HPSG. In Greville Corbett, N. Fraser and S. McGlashan (eds.), *Heads in Grammatical Theory*, pages 186–203, Cambridge: Cambridge University Press.

Copestake, Ann, Flickinger, Daniel P. and Sag, Carl Pollard Ivan A. 2003. Minimal Recursion Semantics. An Introduction.

Fillmore, Charles J., Kay, Paul and O'Connor, Mary Catherine. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language* 64(3), 501–538.

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

Kageyama, Taro. 2001. Word Plus: The Intersection of Words and Phrases. In J. v.d. Weijer and T. Nishihara (eds.), *Issues in Japanese Phonology and Morphology*, pages 245–276, Mouton de Gruyter.

Makino, Seiichi and Tsutsui, Michio. 1986. *A Dictionary of Basic Japanese Grammar*. Tokyo: The Japan Times.

Martin, Samuel E. 1987. *A Reference Grammar of Japanese*. Tokyo: Charles E. Tuttle Company, second edition.

Matsumoto, Y., Kitauchi, A., Yamashita, T., Hirano, Y., Matsuda, H., Takaoka, K. and Asahara, M. 2000. Morphological Analysis System ChaSen verion 2.2.1 Manual.

Oepen, Stephan, Bender, Emily M., Callmeier, Ulrich, Flickinger, Dan and Siegel, Melanie. 2002. Parallel Distributed Grammar Engineering for Practical Applications. In *Proceedings of the Workshop on Grammar Engineering and Evaluation, COLING 2002*, Taipei, Taiwan.

Pollard, Carl and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Studies in Contemporary Linguistics, Chicago, IL and Stanford, CA: The Univeristy of Chicago Press and CSLI Publications.

Radford, Andrew. 1993. Head-Hunting: On the Trail of the Nominal Janus. In Greville Corbett, N. Fraser and S. McGlashan (eds.), *Heads in Grammatical Theory*, pages 73–113, Cambridge: Cambridge University Press.

Sag, Ivan A., Wasow, Thomas and Bender, Emily M. 2003. *Synactic Theory: A Formal Introduction*. Stanford, CA: CSLI, second edition.

Shibatani, Masayoshi and Kageyama, Taro. 1988. Word Formation in a Modular Theory of Grammar: Postsyntactic Compounds in Japanese. *Language* 64, 451–484.

Siegel, Melanie. 1999. The Syntactic Processing of Particles in Japanese Spoken Language. In Jhing-Fa Wang and Chung-Hsien Wu (eds.), *Proceedings of the 13th Pacific Asia Conference on Language, Information and Computation*, Taipei, Taiwan.

Siegel, Melanie. 2000. HPSG Analysis of Japanese. In Wolfgang Wahlster (ed.), *Verbmobil: Foundations of Speech-to-Speech Translation*, Berlin: Springer.

Smith, Jeffrey D. 1999. English Number Names in HPSG. In Gert Webelhuth, Jean-Pierre Koenig and Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, pages 145–160, Stanford, CA: CSLI.

Uszkoreit, Hans, Callmeier, Ulrich, Eisele, Andreas, Schäfer, Ulrich, Siegel, Melanie and Uszkoreit, Jakob. 2004. Hybrid Robust Deep and Shallow Semantic Processing for Creativity Support in Document Production. In *Proceedings of KONVENS 2004*, Vienna, Austria.

Zwicky, Arnold M. 1993. Heads, Bases, and Functors. In Greville C. Corbett, Norman M. Fraser and Scott McGlashan (eds.), *Heads in Grammatical Theory*, pages 292–315, Cambridge: Cambridge University Press.