Implementing the Syntax of Japanese Numeral Classifiers

Emily M. Bender
Department of Linguistics
University of Washington
Box 354340
Seattle WA 98195-4340
ebender@u.washington.edu

Melanie Siegel
Saarland University
Computational Linguistics
PF 15 11 50
D-66041 Saarbrücken
siegel@dfki.uni-sb.de

Abstract

While the sortal constraints associated with Japanese numeral classifiers are well-studied, less attention has been paid to the details of their syntax. We describe an analysis implemented within a broad-coverage HPSG that handles an intricate set of numeral classifier construction types and compositionally relates each to an appropriate semantic representation, using Minimal Recursion Semantics.

1 Introduction

Much attention has been paid to the semantic aspects of Japanese numeral classifiers, and in particular, the semantic constraints which govern which classifiers co-occur with which nouns (Yo, 1993; Bond and Paik, 2000). Here, we focus on a more neglected aspect of this linguistic phenomenon, namely the syntax of numeral classifiers: How they combine with number names to create numeral classifier phrases, how they modify head nouns, and how they can occur as stand-alone NPs. We find that there is both broad similarity and differences in detail across different types of numeral classifiers in their syntactic and semantic behavior. We present semantic representations for two types of numeral classifiers, and describe how they can be constructed compositionally in an implemented broad-coverage HPSG (Pollard and Sag, 1994) for Japanese.

The grammar of Japanese in question is JACY,[1] originally developed as part of the *Verbmobil* project (Siegel, 2000) to handle spoken Japanese, and then extended to handle informal written Japanese (email text; (Siegel and Bender, 2002)) and newspaper text. Recently, it has been adapted to be consistent with the LinGO Grammar Matrix (Bender et al., 2002).

[1] http://www.dfki.uni-sb.de/∼siegel/grammar-download/JACY-grammar.html

2 Types of numeral classifiers

Paik and Bond (2002) divide Japanese numeral classifiers into five major classes: *sortal*, *event*, *mensural*, *group* and *taxanomic*, and several subclasses. The classes and subclasses can be differentiated according to the semantic relationship between the classifiers and the nouns they modify, on two levels: First, what properties of the modified noun motivate the choice of the classifier, and second what properties the classifiers predicate of the nouns. As we are concerned here with the syntax and compositional semantics of numeral classifiers, we will focus only on the latter. Sortal classifiers, (*kind*, *shape*, and *complement* classifiers) serve to individuate the nouns they modify. Event classifiers quantify events, characteristically modifying verbs rather than nouns. Mensural classifiers measure some property of the entity denoted by the noun they modify (e.g., its length). NPs containing group classifiers denote a group or set of individuals belonging to the type denoted by the noun. Finally, taxonomic classifiers force a kind or species reading on an NP.

In this paper, we will treat the syntax and compositional semantics of sortal and mensural classifiers. However, we believe that our general analysis can be extended to treat the full range of classifiers in Japanese and similar languages.

3 Data: Constructions

Internally, Japanese numeral classifier expressions consist of a number name followed by a numeral classifier (1a,b,c). In this, they resemble date expressions (1d).[2]

(1) a. juu mai
10 NumCl

b. juu en
10 yen

c. juu kagetsu
10 month
'10 months'

d. juu kagetsu
10 month
'October'

In fact, both numeral classifiers and date expressions are tagged as numeral classifiers by the morphological analyzer ChaSen (Asahara and Matsumoto, 2000). However, date expressions do not have the same combinatoric potential (syntactic or semantic) as numeral classifiers. We thus give date expressions a distinct analysis, which we will not describe here.

Externally, numeral classifier phrases (NumClPs) appear in at least four different contexts: alone, as anaphoric NPs (2a); preceding a head noun, linked by the particle *no* (2b); immediately following a head noun (2c); and 'floated', right after the associated noun's case particle or right before the verb (2d). These constructions are distinguished pragmatically (Downing, 1996).[3]

(2) a. ni hiki wo kau
2 NumCl ACC raise
'(I) am raising two (small animals).'

b. ni hiki no neko wo kau
2 NumCl GEN cat ACC raise
'(I) am raising two cats.'

c. neko ni hiki wo kau
cat 2 NumCl ACC raise
'(I) am raising two cats.'

d. neko wo (ni hiki) ie de
cat ACC (2 NumCl) house LOC
(ni hiki) kau
(2 NumCl) raise
'(I) am raising two cats in my house.'

NumClPs can be modified by elements such as *yaku* 'approximately' (before the number name) or *mo* 'even' (after the floated numeral classifiers).

The above examples illustrate the contexts with a sortal numeral classifier, but mensural numeral classifiers can also appear both as modifiers (3a) and as NPs in their own right (3b):

(3) a. ni kiro no ringo wo katta
2 NumCl (kg) GEN apple ACC bought
'(I) bought two kilograms of apples.'

b. ni kiro wo katta
2 NumCl (kg) ACC bought
'(I) bought two kilograms.'

NumClPs serving as NPs can also appear as modifiers of other nouns:

(4) a. san nin no deai wa 80 nen haru
3 NumCl GEN meeting TOP 80 year spring
'The three's meeting was in the spring of '80.'

b. ichi kiro no nedan ha hyaku en desu
1 kg GEN price TOP 100 yen COPULA
'The price of/for 1 kg is 100 yen.'

As a result, tokens following the syntactic pattern of (2b) and (3a) are systematically ambiguous, although the non-anaphoric reading tends to be preferred.

Certain mensural classifiers can be followed by the word *han* 'half':

(5) ni kiro han
two kg half
'two and a half kilograms'

In order to build their semantic representations compositionally, we make the numeral classifier (here, *kiro*) the head of the whole expression, and *ni* and *han* its dependents. *Kiro* can then orchestrate the semantic composition of the two dependents as well as the composition of the whole expression with the noun it modifies (see §6 below).

Although they aren't tagged as numeral classifiers by ChaSen, we extended our analysis of mensural classifiers to certain elements that appear before numbers, namely currency symbols (such as *$*), and prefixes like *No.* 'number' in (6).

[2] Note that many of the time units are ambiguous with date expressions, although some, like the one for months shown in (1), are distinguished.

[3] Downing also notes NumClPs following the head noun with an intervening *no*. As this rare construction did not appear in our data, we have not incorporated it into our account.

(6) kouza No. 1234 gou
account number 1234 number
'account number 1234'

Finally, we found that number names can sometimes occur without numeral classifiers, either as modifiers of nouns or as anaphora:

(7) (kouza) 1234 wo tojitai
(account) 1234 ACC close.volitional
'(I) want to close (account) 1234.'

Due to space considerations, we won't describe our analysis of such bare number names here.

4 Data: Distribution

We used ChaSen to segment and tag 10,000 paragraphs of the Mainichi Shinbun 2002 corpus. Of the resulting 490,202 words, 11,515 (2.35%) were tagged as numeral classifiers. 4,543 of those were potentially time/date expressions, leaving 6,972 numeral classifiers, or 1.42% of the words. 203 orthographically distinct numeral classifiers occur in the corpus. The most frequent is *nin* (the numeral classifier for people) which occurs 1,675 times.

We sampled 100 sentences tagged as containing numeral classifiers to examine the distribution of the constructions outlined in §3. These sentences contained a total of 159 numeral classifier phrases and the vast majority (128) were stand-alone NPs. This contrasts with Downing's (1996) study of 500 examples from modern works of fiction and spoken texts, where most of the occurrences are not anaphoric. Furthermore, while our sample contains no examples of the floated variety, Downing's contains 96. The discrepancy probably arises because Downing only included sortal numeral classifiers, and not any other type. Another possible contributing factor is the effect of genre. In future work we hope to study the distribution of both the types of classifiers and the constructions involving them in the Hinoki treebank (Bond et al., 2004).

5 Semantic Representations

One of our main goals in implementing a syntactic analysis of numeral classifiers is to compositionally construct semantic representations, and in particular, Minimal Recursion Semantics (MRS) representations (Copestake et al., 2003; Copestake et al., 2001). Abstracting away from handle constraints,[4] illocutionary force, tense/aspect, and the unexpressed subject, the representation we build for (2b,c) is as in (8).[5]

(8) **_cat_n_rel**(x), **udef_rel**(x), **card_rel**(x,"2"),
_raise_v_rel(z,x)

This can be read as follows: A relation of raising holds between z (the unexpressed subject), and x. x denotes a cat entity, and is bound by an underspecified quantifier (as there is no explicit determiner). x is also an argument of a **card_rel** (short for 'cardinal_relation'), whose other argument is the constant value 2, meaning that there are in fact two cats being referred to.[6]

For anaphoric numeral classifiers, the representation contains an underspecified **noun_relation**, to be resolved in further processing to a specific relation:

(9) **noun_relation**(x), **udef_rel**(x), **card_rel**(x,"2"),
_raise_v_rel(z,x)

Mensural classifiers have somewhat more elaborated semantic representations, which we treat as similar to English measure NPs (Flickinger and Bond, 2003). On this analysis, the NumClP denotes the extent of some dimension or property of the modified N. This dimension or property is represented with an underspecified relation (**unspec_adj_rel**), and a **degree_rel** relates the measured amount to the underspecified adjective relation.[7] The underspecified adjective relation modifies the N in the usual way. This is illustrated in (10), which is the semantic representation assigned to (3a).[8]

[4]The potentially underspecified MRS representation of scope.

[5]By convention, the predicate names for lexically contributed relations reflect the orthography of the lexical items that introduce them. In this paper, we are using English translations of the predicate names for expository convenience.

[6]We take it as implicit in this representation that uncountable nouns are individuated when they appear as arguments of a **card_rel**.

[7]For clarity, we show the relation between the **degree_rel** and the measure phrase by giving the index of the measure phrase a role in the **degree_rel**. In the current implementation, however, this relationship is represented with identity of handles (see (19)).

[8]The relationship between the **degree_rel** and the **unspec_adj_rel** is not entirely apparent in this abbreviated notation. The first argument of the **degree_rel** is in fact the predicate name of the **unspec_adj_rel**, and not the whole relation.

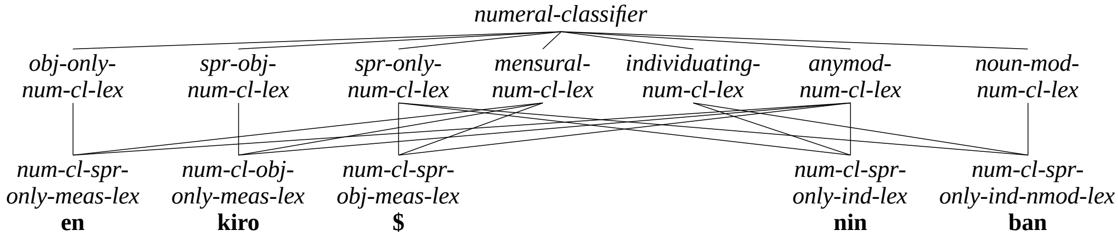


Figure 1: Type hierarchy under *numeral-classifier*

(10) **_kilogram_n_rel**(x), **udef_rel**(x), **card_rel**(x,"2"), **degree_rel**(**unspec_adj_rel**, x), **unspec_adj_rel**(y), **_apple_n_rel**(y), **udef_rel**(y), **_buy_v_rel**(z,y)

When mensural NumClPs are used anaphorically (3b), the element modified by the **_unspec_adj_rel** is an underspecified **_noun_relation**, analogously to the case of sortal NumClPs used anaphorically:

(11) **_kilogram_n_rel**(x), **udef_rel**(x), **card_rel**(x,"2"), **degree_rel**(**unspec_adj_rel**, x), **unspec_adj_rel**(y), **noun_relation**(y), **udef_rel**(y), **_buy_v_rel**(z,y)

6 Implementing an Analysis

Our analysis consists of: (1) a lexical type hierarchy cross-classifying numeral classifiers along three dimensions (Fig 1), (2) a special lexical entry for *no* for linking NumClPs with nouns, (3) a unary-branching phrase structure rules for promoting NumClPs to nominal constituents.

6.1 Lexical types

Fig 1 shows the lexical types for numeral classifiers, which are cross-classified along three dimensions: semantic relationship to the modified noun (*individuating* or *mensural*), modificational possibilities (NPs or PPs: *anymod*/NPs: *noun-mod*), and relationship to the number name (number name precedes: *spr-only*, number name precedes but may take *han*: *spr-obj*, number name follows: *obj-only*). Not all the possibilities in this space are instantiated (e.g., we have found no sortal classifiers which can take *han*), but we leave open the possibility that we may find in future work examples that fill in the range of possibilities.

The constraint in (12) ensures that all numeral classifiers have the head type *num-cl_head*, as required by the unary phrase structure rule discussed in §6.4 below. Furthermore, it identifies two key pieces of semantic information made available for further composition, the INDEX and LTOP (local top handle) of the modified element with the numeral classifier's own INDEX and LTOP, as these are intersective modifiers (Bender et al., 2002). The constraints on the type *num-cl_head* (not shown here) ensure that numeral classifiers can modify only saturated NPs or PPs (i.e., NPs marked with a case postposition *wo* or *ga*), and that they only combine via intersective head-modifier rules.[9]

(12) *numeral-classifier* :=

$$\left[\begin{array}{ll} \text{...CAT.HEAD} & \left[\begin{array}{l} \textit{num-cl\_head} \\ \text{MOD} \left\langle \left[\begin{array}{ll} \text{...INDEX} & \boxed{1} \\ \text{...LTOP} & \boxed{2} \end{array}\right] \right\rangle \end{array}\right] \\ \text{...CONT.HOOK} & \left[\begin{array}{ll} \text{INDEX} & \boxed{1} \\ \text{LTOP} & \boxed{2} \end{array}\right] \end{array}\right]$$

The constraints on the types *spr-only-num-cl-lex*, *obj-only-num-cl-lex* and *spr-obj-num-cl-lex* account for the position of the numeral classifier with respect to the number name and for the potential presence of *han*. Both the number name (a phrase of head type *int_head*) and *han* (given the distinguished head value *han_head*) are treated as dependents of the numeral classifier expression, but variously as specifiers or complements according to the type. In the JACY grammar, specifiers immediately precede their heads, while complements are not required to do so and can even follow their heads (in rare cases). Given all this, in the ordinary case (*spr-only-num-cl-lex*), we treat the number name as the specifier of the numeral classifier. The other two cases involve numeral classifiers taking complements: with no specifier, in the case of pre-number unit expressions like the symbol *$* (*obj-only-num-cl-lex*) and both a

[9]Here and throughout, we have suppressed certain details of the feature structures and abbreviated feature paths. Angle brackets with exclamation points inside (⟨! !⟩) indicate difference lists, used to enable list appends in unifi cation.

number-name specifier and the complement *han* in the case of unit expressions appearing with *han* (*spr-obj-num-cl-lex*).[10] Finally, the type *spr-obj-num-cl-lex* does some semantic work as well, providing the **plus_rel** which relates the value of the number name to the "$\frac{1}{2}$" contributed by *han*, and identifying the ARG1 of the **plus_rel** with the XARG the SPR and COMPS so that they will all share an index argument (eventually the index of the modified noun for sortal classifiers and of the measure noun relation for mensural classifiers). The constraints which implement these aspects of our analysis are sketched in (13)–(15).

(13) *spr-only-num-cl-lex* :=

$$\begin{bmatrix} \text{...VAL} & \begin{bmatrix} \text{SUBJ} & \textit{null} \\ \text{OBJ} & \textit{null} \\ \text{SPR} & \begin{bmatrix}\text{...CAT.HEAD} & \textit{int\_head}\end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(14) *obj-only-num-cl-lex* :=

$$\begin{bmatrix} \text{...VAL} & \begin{bmatrix} \text{SUBJ} & \textit{null} \\ \text{OBJ} & \begin{bmatrix}\text{...CAT.HEAD} & \textit{int\_head}\end{bmatrix} \\ \text{SPR} & \textit{null} \end{bmatrix} \end{bmatrix}$$

(15) *spr-obj-num-cl-lex* :=

$$\begin{bmatrix} \text{...VAL} & \begin{bmatrix} \text{SUBJ} & \textit{null} \\ \text{OBJ} & \begin{bmatrix}\text{...CAT.HEAD} & \textit{han\_head} \\ \text{...CONT.HOOK} & \begin{bmatrix}\text{LTOP} & \boxed{1} \\ \text{XARG} & \boxed{2}\end{bmatrix}\end{bmatrix} \\ \text{SPR} & \begin{bmatrix}\text{...CAT.HEAD} & \textit{int\_head} \\ \text{...CONT.HOOK} & \begin{bmatrix}\text{LTOP} & \boxed{3} \\ \text{XARG} & \boxed{2}\end{bmatrix}\end{bmatrix} \end{bmatrix} \\ \text{...RELS} & \langle ! \begin{bmatrix}\textit{plus-relation} \\ \text{ARG1} & \boxed{2} \\ \text{TERM1} & \boxed{3} \\ \text{TERM2} & \boxed{1}\end{bmatrix} ! \rangle \end{bmatrix}$$

In the second dimension of the cross-classification, *anymod-num-cl-lex* and *noun-mod-num-cl-lex* constrain what the numeral classifier may modify, via the MOD value.

When numeral classifiers appear before the head noun, they are linked to it with *no*, which mediates the modifier-modifiee relationship (see (2) and §6.2). However, numeral classifiers can appear after the noun (2c), modifying it directly. Some numeral classifiers can also 'float' outside the NP, either immediately after the case postposition or to the position before the verb (2d).[11] While we leave the latter kind of float to future work (see §7), we handle the former by allowing most numeral classifiers to appear as post-head modifiers of PPs. Thus *noun-mod-num-cl-lex* further constrains the HEAD value of the element on the MOD list to be *noun_head*, but *anymod-num-cl-lex* leaves it as inherited (*noun-or-case-p_head*). This type does, however, constrain the modifier to show up after the head ([POSTHEAD *right*]), and further constrains the modified head to be [NUCL *nucl_plus*], in order to rule out vacuous attachment ambiguities between numeral classifiers attaching to the right and other modifiers appearing to the left of the NP.

(16) *noun-mod-num-cl-lex* :=

$$\begin{bmatrix}\text{...MOD} \ \langle \begin{bmatrix}\text{...HEAD } \textit{noun\_head}\end{bmatrix} \rangle\end{bmatrix}$$

(17) *anymod-num-cl-lex* :=

$$\begin{bmatrix} \text{...HEAD} & \begin{bmatrix}\text{MOD} \ \langle \begin{bmatrix}\text{LOCAL.NUCL } \textit{nucl\_plus}\end{bmatrix} \rangle \\ \text{POSTHEAD } \textit{right}\end{bmatrix}\end{bmatrix}$$

The final dimension of the classification captures the semantic differences between sortal and mensural numeral classifiers. The sortal numeral classifiers contribute no semantic content of their own.[12] They are therefore constrained to have empty RELS and HCONS lists:

(18) *individuating-num-cl-lex* :=

$$\begin{bmatrix} \text{...CONT} & \begin{bmatrix}\text{RELS} & \langle ! \ ! \rangle \\ \text{HCONS} & \langle ! \ ! \rangle\end{bmatrix}\end{bmatrix}$$

In contrast, mensural numeral classifiers contribute quite a bit of semantic information, and therefore have quite rich RELS and HCONS values. As shown in (19), the *noun-relation* is identified with the lexical key relation value (LKEYS.KEYREL) so

[10] Because numeral classifiers are analyzed as taking post-head complements in these two cases, the head type *num-cl_head* is a subtype of *init-head*, which contrasts with *final-head*. These types are used by the head-complement rules to determine the order of the head and complements.

[11] Those that can't include expressions like *gou* in (i), cf. (ii):

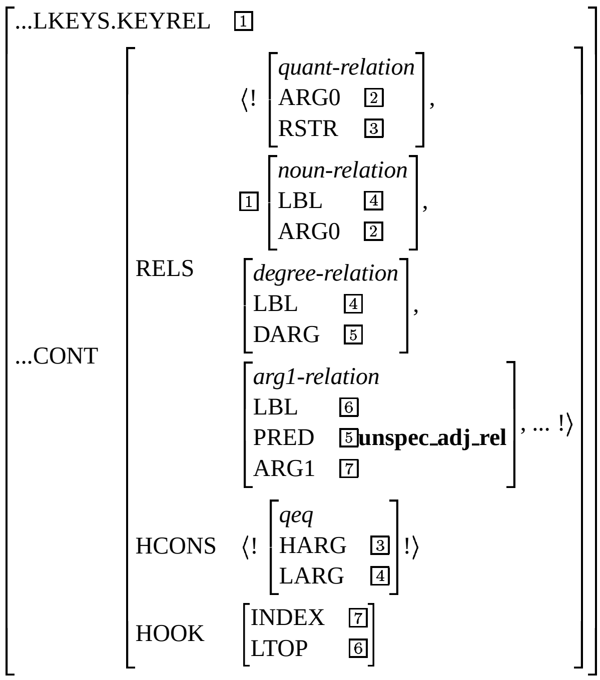
(i) kouza 1234 gou wo tojitai
account 1234 number ACC close.volitional
'(I) want to close account number 1234.'

(ii) *kouza wo 1234 gou tojitai

[12] The individuating function they serve we take to be implicit in the linkage they provide between the **card_rel** and the noun relation. See note 6.

that specific lexical entries of this type can easily further specify it (e.g., *kiro* constraints its PRED to be **_kilogram_n_rel**). The type also makes reference to the HOOK value so that the INDEX and LTOP (also the INDEX and LTOP of the modified noun, see (12)) can be identified with the appropriate values inside the RELS list. The length of the RELS list is left unbounded, because some mensural classifiers also inherit from *spr-obj-num-cl-lex*, and therefore must be able to add the **plus_rel** to the list.

(19) *mensural-num-cl-lex* :=

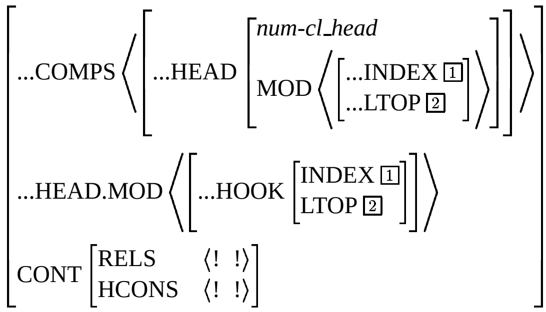


The types in the bottom part of the hierarchy in Fig 1 join the dimensions of classification. They also do a little semantic work, making the INDEX and LTOP of the modified noun available to their number name argument, and, in the case of subtypes of *mensural-num-cl-lex*, they constrain the final length of the RELS list, as appropriate.

6.2 The linker *no*

We posit a special lexical entry for *no* which mediates the relationship between NumClPs and the nouns they modify. In addition to the constraints that it shares with other entries for *no* and other modifier-heading postpositions, this special *no* is subject to the constraints shown in (20). These specify that *no* makes no semantic contribution, that it takes a NumClP as a complement, and that the element on the MOD list of *no* shares its local top handle and index with the element on the MOD list of the NumClP (i.e., that *no* effectively inherits its complement's MOD possibility). Even though (most) numeral classifiers can either modify NPs or PPs, all entries for *no* are independently constrained to only modify NPs, and only as pre-head modifiers.

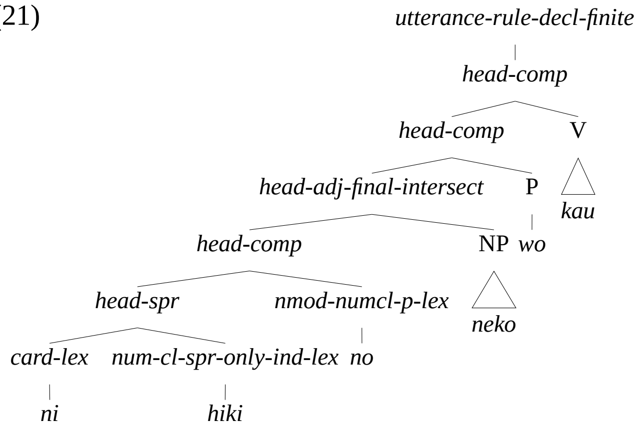
(20) *nmod-numcl-p-lex* :=



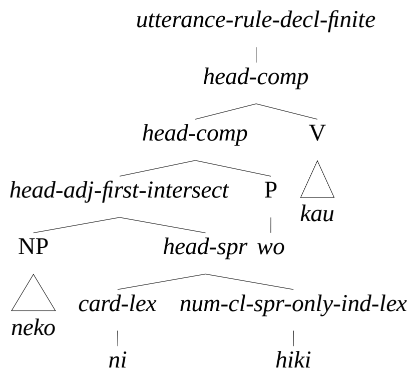
6.3 Examples: NumClPs as Modifiers

We illustrate our analysis with sample derivations, displayed as trees with (abbreviated) rule names and lexical types on the nodes. (21) corresponds to (2b), (22) to (2c), and (23) to a shortened (2d).

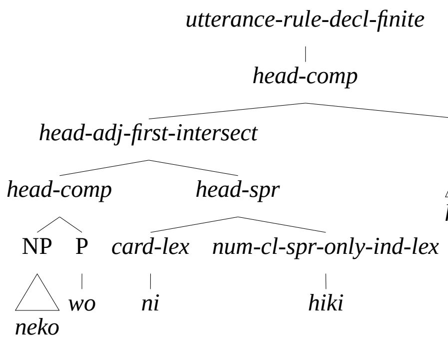
(21)



(22)



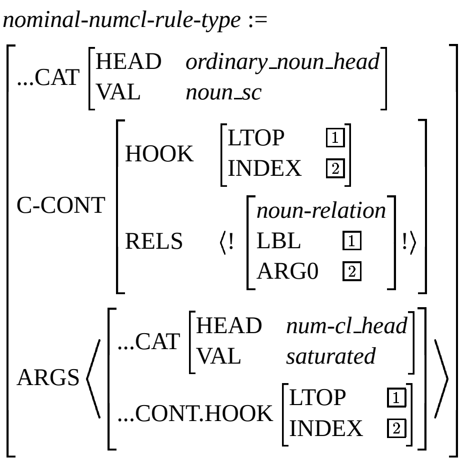
(23)



6.4 Unary-branching phrase structure rule

We treat NumClPs serving as nominal constituents by means of an exocentric unary-branching rule.[13] This rule specifies that the mother is a noun subcategorized for a determiner specifier (these constraints are expressed on *noun_sc*), while the daughter is a numeral classifier phrase whose valence is saturated. Furthermore, it contributes (via its C-CONT, or constructional content feature) an underspecified *noun-relation* which serves as the thing (semantically) modified by the numeral classifier phrase. The reentrancies required to represent this modification are implemented via the LTOP and INDEX features.

(24)



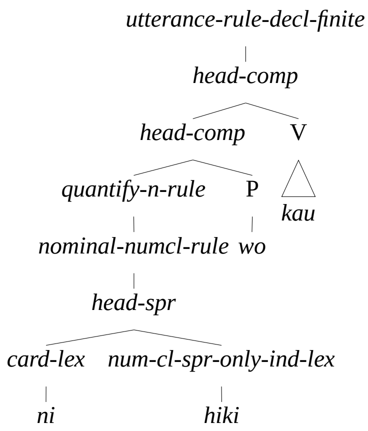
This rule works for both sortal and mensural NumClPs, as both are expecting to modify a noun.

6.5 Examples: NumClPs as Nouns

Again, we illustrate the interaction of these various constraints with an example derivation (25) for (2a).

[13]In the analysis of number names used as NumClPs, we posit a second unary-branching rule. The mother of that rule (a NumClP) can then serve as the daughter of the rule discussed here.

(25)



7 Future Work

We have not yet implemented an analysis of preverbal floated NumClPs, but we sketch one here. The key is that NumClPs are treated as simple modifiers, not quantifiers. Therefore, they can attach syntactically to the verb, but semantically to one of its arguments. In our HPSG analysis, the verb will have unsaturated valence features, making the indices of its arguments 'visible' to any modifiers attaching to it.

There appear to be constraints on which arguments can 'launch' floated quantifiers, although their exact nature is as yet unclear. Proposals include: only nominals marked with the case particles *ga* or *wo* (Shibatani, 1978), only subjects or direct objects (Inoue, 1978), or c-command-based constraints (Miyagawa, 1989). While there are exceptions to all of these generalizations, Downing (1996) notes that the vast majority of actually occurring cases satisfy all of them, and further that it is primarily *intransitive* subjects which participate in the construction.

These observations will help considerably in reducing the ambiguity inherent in introducing an analysis of floated NumClPs. We could constrain floated NumClPs to only modify intransitive verbs (semantically modifying the subject) or transitive verbs (semantically modifying the object). Some ambiguity will remain, however, as the pre-verbal and post-nominal positions often coincide.

Also missing from our analysis are the sortal constraints imposed by classifiers on the nouns they modify. In future work, we hope to merge this analysis with an implementation of the sortal constraints, such as that of Bond and Paik (2000) . We believe that such a merger would be extremely useful: First, the sortal constraints could be used to narrow down the possible referents of anaphoric uses of NumClPs. Second, sortal constraints could reduce ambiguity in NumClP+*no*+N strings, whenever they could rule out the ordinary numeral classifier use, leaving the anaphoric interpretation (see (4) above). Third, sortal constraints will be crucial in generation (Bond and Paik, 2000). Without them, we would propose an additional string for each sortal classifier whenever a **card_rel** appears in the input semantics, most of which would in fact be unacceptable. Implementing sortal constraints could be simpler for generation than for parsing, since we wouldn't need to deal with varying inventories or metaphorical extensions.

8 Conclusion

Precision grammars require compositional semantics. We have described an approach to the syntax of Japanese numeral classifiers which allows us to build semantic representations for strings containing these prevalent elements – representations suitable for applications requiring natural language understanding, such as (semantic) machine translation and automated email response.

Acknowledgements

This research was carried out as part a joint R&D effort between YY Technologies and DFKI, and we are grateful to both for the opportunity. We would also like to thank Francis Bond, Dan Flickinger, Stephan Oepen, Atsuko Shimada and Tim Baldwin for helpful feedback in the process of developing and implementing this analysis and Setsuko Shirai for grammaticality judgments. This research was partly supported by the EU project DeepThought IST-2001-37836.

References

Masayuki Asahara and Yuji Matsumoto. 2000. Extended models and tools for high-performance part-of-speech tagger. In *Coling 2000*, Saarbrücken, Germany.

Emily M. Bender, Dan Flickinger, and Stephan Oepen. 2002. The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Proceedings of the Workshop on Grammar Engineering and Evaluation, Coling 2002*, pages 8–14, Taipei.

Francis Bond and Kyoung-Hee Paik. 2000. Reusing an ontology to generate numeral classifiers. In *Coling 2000*, Saarbrücken, Germany.

Francis Bond, Sanae Fujita, Chikara Hashimoto, Kaname Kasahara, Shigeko Nariyama, Eric Nichols, Akira Ohtani, Takaaki Tanaka, and Shigeaki Amano. 2004. The Hinoki Treebank: A treebank for text understanding. In *IJC-NLP-2004*.

Ann Copestake, Alex Lascarides, and Dan Flickinger. 2001. An algebra for semantic construction in constraint-based grammars. In *ACL 2001*, Toulouse, France.

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

Pamela Downing. 1996. *Numeral Classifier Systems: The Case of Japanese*. John Benjamins, Philadelphia.

Dan Flickinger and Francis Bond. 2003. A two-rule analysis of measure noun phrases. In Stefan Müller, editor, *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar*, pages 111–121, Stanford CA. CSLI Publications.

Kazuko Inoue. 1978. *Nihongo no Bunpou Housoku*. Tasishuukan, Tokyo.

Shigeru Miyagawa. 1989. *Structure and Case Marking in Japanese*. Academic Press, New York.

Kyounghee Paik and Francis Bond. 2002. Spatial representation and shape classifiers in Japanese and Korean. In David I. Beaver, Luis D. Casillas Martínez, Brady Z. Clark, and Stefan Kaufmann, editors, *The Construction of Meaning*, pages 163–180. CSLI Publications, Stanford CA.

Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. U of Chicago Press, Chicago.

Masayoshi Shibatani. 1978. *Nihongo no Bunseki*. Tasishuukan, Tokyo.

Melanie Siegel and Emily M. Bender. 2002. Efficient deep processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and Standardization, Coling 2002*, Taipei.

Melanie Siegel. 2000. HPSG analysis of Japanese. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer, Berlin.

Matsumoto Yo. 1993. Japanese numeral classifiers: A study of semantic categories and lexical organization. *Linguistics*, 31:667–713.