

Parsed Corpus as a source for testing generalizations in Japanese syntax

Kishimoto, Hideki and Prashant Pardeshi

Kobe University and NINJAL

Japanese is a language with SOV order, but it permits word order variation in the presence of a reordering operation, which is generally referred to as scrambling. Unmarked transitive clauses have nominative subjects and accusative objects. While the predicate is placed in final position, the order of subjects and objects in transitive predicates are relatively free, because they can be scrambled. Thus, we can have the two different arrangements; one is ‘nominative-accusative’ order and the other ‘accusative-nominative’ order, as in (1).

- (1) a. John-ga Mary-o home-ta.
 John-NOM Mary-ACC admire-PAST
 ‘John admired Mary.’
- b. Mary-o John-ga home-ta.
 Mary-ACC John-NOM admire-PAST
 ‘John admired Mary.’

While native speakers generally agree that (1a) represents the basic word order of Japanese, and (1b) is a derived one. Note that scrambling does not induce much of semantic effects, but still, it is imaginable that the derived order is generated for good reasons, i.e. with some motivations. Given that the basis word order is unmarked and neutral, the derived word represented in (1b) is expected to have much less frequency. The word order variation can be easily checked in parsed corpora, and can find a significantly large number of occurrences of sentences with SOV word order, when compared with the occurrences of the sentences with OSV word order in which the subject as well as the object is overt (1941 versus 97 tokens). The corpus data provide one handy way of confirming the native speaker’s intuition about the basic word order and is also useful when teaching an elementary course in Japanese syntax to non-native speakers.

More or less the same applies to the ordering of dative and nominative arguments of some transitive predicates. Here, we need to pay attention to individual cases, because these two patterns could be basic. Japanese can have nominative-dative patterns for some non-stative predicates, as in (2a) whereas the reversed case-marking pattern is obtained when the predicates are stative, as in (2b).

- (2) a. John-ga Mary-ni at-ta.
 John-NOM Mary-DAT meet-PAST
 ‘John met with Mary.’

- b. John-ni kodomo-ga i-ru.
 John-DAT child-NOM be-PRES
 ‘John has a brother.’

A corpus search shows 33 tokens for stative predicates taking dative-nominative order, and 3 tokens for non-stative predicates taking nominative-dative order. Even though the limited number of tokens are found by the search, this result, which gives only the basic word order, would be naturally expected that if the occurrence of scrambled order is very restricted in naturally-occurring corpus.

Parsed corpora also provide some clues to an on-going issue on the order of the two internal arguments of ditransitive predicates, which express the meaning of transfer from one place to the other. Ditransitive predicates select dative-marked goal and accusative-marked theme arguments, alongside nominative subjects.

- (3) John-ga Mary-ni hon-o age-ta.
 John-NOM Mary-DAT book-ACC give-PAST
 ‘John gave Mary a present.’

With regard to the word order of the internal arguments, some researchers claim that the dative-accusative order is basic, but others claim that the accusative-dative order is basic (e.g. Hoji 1985, Miyagawa and Tsujioka 2004). Native speakers are likely to judge the dative-accusative order as basic, but their judgments are not really as robust as other cases. Moreover, a number of more subtle conflicting data are discussed in the literature—one major reason why the conflicting claims are available.

Naturally occurring examples in our parsed corpus provide support for the former view taking the dative-accusative order to be basic, because we can find by far more examples with the dative-accusative order than those with the accusative-order (34 versus 16 tokens). Note, however, that the percentage of accusative-dative order (relative to dative-accusative order) is much higher than that of accusative-nominative order (relative to nominative-accusative order). This can be taken as a reflection of the fact that native speakers are often uncertain about which order is basic in the case of ditransitive verbs.

Finally, one of the cautions about the search results is that they may also contain structures that the user is not looking for or interested in. For instance, there are many kinds of *ni* in Japanese, and it is controversial whether *ni* is dative case, a form of copula or something else. The categories employed in annotation do not always correspond to the expectations of the user. The user therefore needs to carefully choose only those tokens that he/she is interested in manually, however, this task is far more manageable than searching manually without using a parsed corpus.

References

- Hoji, Hajime. 1985. *Logical Form Constraints and Configurational Structures in Japanese*. Doctoral dissertation, University of Washington, Seattle.
- Miyagawa, Shigeru, and Takae Tsujioka. 2004. Argument structure and ditransitive verbs in Japanese. *Journal of East Asian Linguistics* 13. 1-38.