#### LiLT volume 18, issue 2

#### July 2019

## Parsed Corpus as a Source for Testing Generalizations in Japanese Syntax<sup>\*</sup>

HIDEKI KISHIMOTO, Kobe University PRASHANT PARDESHI, National Institute for Japanese Language and Linguistics (NINJAL)

#### Abstract

In this paper, we discuss constituent ordering generalizations in Japanese. Japanese has SOV as its basic order, but a significant range of argument order variations brought about by 'scrambling' is permitted. Although scrambling does not induce much in the way of semantic effects, it is conceivable that marked orders are derived from the unmarked order under some pragmatic or other motivations. The difference in the effect of basic and derived order is not reflected in native speaker's grammaticality judgments, but we suggest that the intuition about the ordering of arguments may be attested in corpus data. By using the Keyaki treebank (a proper subset of which is NINJAL Parsed Corpus of Modern Japanese (NPCMJ)), it is shown that the naturallyoccurring corpus data confirm that marked orderings of arguments are less frequent than their unmarked ordering counterparts. We suggest some possible motivations lying behind the argument order variations.

<sup>\*</sup> An earlier version of this paper was presented at the NINJAL Internal Symposium 'Exploiting Parsed Corpora: Application in Research, Pedagogy, and Processing' held at the National Institute for Japanese Language and Linguistics (December 9-10). We are grateful to Beatrice Santorini, Takehiko Maruyama, Koji Mineshima, Yusuke Kubota, Iku Nagasaki, and participants of the conference for various comments. We are also thankful to Stephen Horn, Alastair Butler, and Ayaka Suzuki for their generous assistance on searching the Keyaki treebank and NINJAL Parsed Corpus of Modern Japanese (NPCMJ). We are grateful to the two reviewers for

#### 1 Introduction

In this paper, we will discuss constituent ordering generalizations in Japanese. Japanese has SOV as its basic order, but a significant range of word order variation is permitted. In Japanese, reordering of arguments can be brought about by a syntactic operation of 'scrambling'. Unmarked transitive clauses have nominative subjects followed by accusative objects. While the predicate is consistently placed in final position, the order of arguments is relatively free because they can be scrambled. Simple transitive clauses can have two different arrangements of nominative and accusative arguments; namely, 'nominative-accusative' and 'accusative-nominative' order.

Scrambling does not induce much in the way of semantic effects, and it is sometimes claimed that scrambling can apply freely. Nevertheless, native speakers generally agree that the 'nominative-accusative' order represents the basic word order of Japanese transitive clauses, while 'accusative-nominative' order is derived. (Note that since scrambling does not change grammatical relations or logical meanings of sentences, it is fair to say that pre-scrambling structures represent ones which are virtually identical to surface forms.<sup>1</sup>) It is conceivable that the derived order is generated under some motivation, e.g. a long, complex object is positioned before a short subject to facilitate the processing of the sentence. The difference in the effect of basic and derived order is not reflected in the native speaker's grammaticality judgments, but we suggest that the intuition about the ordering of arguments may be attested in corpus data. By using two parsed corpora, viz. Keyaki treebank and NINJAL Parsed Corpus of Modern Japanese (NPCMJ),<sup>2</sup> we will argue that the more marked the ordering of arguments is, the less frequently it is attested in the naturally-occurring corpora.

The discussion in this paper proceeds as follows. In section 2, we will discuss how arguments are ordered in Japanese and review some proposals in the literature discussing how basic as well as derived word order is determined in Japanese. Following this, we will present a

their comments and suggestions. The research reported in this paper was supported by the JSPS KAKENHI (Grant Numbers JP16K02628 and JP15H03210) and NIN-JAL Institute-based Project "Development of and Linguistic Research with a Parsed Corpus of Japanese" (PI: Prashant Pardeshi).

<sup>&</sup>lt;sup>1</sup>The term surface form roughly corresponds to S-structure (the term countenanced in generative grammar). Here, we are using the term in a pre-theoretical sense. Needless to say, all of the discussions presented in the paper are concerned with the 'shallow' ordering generalizations.

<sup>&</sup>lt;sup>2</sup>The NPCMJ corpus is a proper subset of the Keyaki treebank. However, Keyaki treebank and NPCMJ differ in terms of the amount of information added in annotation and also the format used.

Parsed Corpus for Testing Generalizations in Japanese Syntax / 3  $\,$ 

case study on the ordering of arguments, including canonical and noncanonical transitive clauses and ditransitive clauses in section 3. We will present a summary of our findings in section 4.

### 2 The ordering of arguments and markedness

Japanese is a language with SOV word order but allows arguments to switch their positions rather flexibly. Variation in argument order is permitted owing to a reordering operation, which is generally referred to as scrambling.<sup>3</sup> In Japanese, transitive clauses have the basic word order in which nominative subjects precedes accusative objects, as exemplified in (1a).

- (1) a. John-ga Mary-o home-ta. John-NOM Mary-ACC praise-PAST 'John praised Mary.'
  - b. Mary-o John-ga home-ta. Mary-ACC John-NOM praise-PAST 'John praised Mary.'

The predicate has a fixed position; it invariantly appears in clausefinal position. The order of subjects and objects in transitive predicates is relatively free, since they can be scrambled. Thus, in transitive clauses, we can have two different arrangements of the arguments; one is 'nominative-accusative' order, as in (1a), and the other 'accusativenominative' order, as in (1b).

In canonical transitive clauses, the speakers' intuition is relatively clear, and it is generally agreed that the transitive clause in (1a) represents the basic word order, and (1a) is a non-basic one, derived by scrambling. For obvious reasons, this intuition is not reflected in grammaticality judgments; the basic and the derived sentences in (1) are both acceptable. However, we will suggest that the intuition on the basic and derived word order is reflected in the number of tokens found in naturally-occurring corpus data, in that the corpus data show that nominative subjects predominately precede accusative objects, as predicted by markedness theory (see Tomic 1989). In non-canonical case-marking clauses, there arises an

<sup>&</sup>lt;sup>3</sup>Terms such as 'scrambling' and 'reordering operation' are meant to be descriptive labels and are used in a theory-neutral manner. In generative grammar, scrambling is treated as a movement transformation operation, but there are other frameworks that do not employ transformations for descriptions of 'displacement' phenomena. We are not committed to making any theoretical claim here.

issue over the basic order of arguments, because a variety of casemarking patterns obtains. In Japanese, there are at least three types of non-canonical transitive clauses, namely, transitive clauses with dative.subject-nominative.object, nominative.subject-dative.object, and nominative.subject-nominative.object alignments. We suggest that the same point can be made regarding the basic word order of ditransitive clauses, namely, clauses with basic word order are found more frequently than clauses with derived word order in corpus data.

With ditransitive verbs, which take two internal arguments, i.e. dative and accusative objects, speakers' intuition about their basic word order is not so clear, and it is not so easy to judge which order represents the basic one, although native speakers tend to take the 'dativeaccusative' rather than 'accusative-dative' to be basic. It is suggested that this kind of uncertainty about word order variation is also reflected in the token frequency of the ditransitive examples in the corpora.

### 3 Testing word order generalizations

In this section, we will review some arguments for determining the basic word order in Japanese and then present a case study of how some generalizations on the ordering of arguments can be attested by using two Japanese corpora: Keyaki treebank and NPCMJ.

#### 3.1 Canonical transitive clauses

As noted at the beginning, the basic word order of a canonical transitive clause is relatively uncontroversial since naive speakers' judgments are fairly clear. Nevertheless, native speakers' naïve judgments on the ordering of arguments is not necessarily reliable, hence more empirical demonstration of the basic word ordering of arguments is needed. In grammatical studies of Japanese, there are at least two arguments offered for demonstrating that 'nominative-accusative' rather than 'accusative-nominative' is the basic word order pattern in canonical transitive clauses. In what follows, we will discuss the cases of numeral quantifier floating and quantifier scope, the syntactic behavior of which suggests that the 'nominative-accusative' ordering is basic, and hence 'accusative-nominative' ordering is derived. <sup>4</sup>

One argument in support of the view that 'nominative-accusative' ordering is basic may be derived from the facts of numeral quantifier floating (NQ floating) discussed by Kuroda (1983). In Japanese,

 $<sup>^4</sup>$ The discussions of numeral quantifier floating and quantifier scope are based on the classic baseline arguments. We are presenting them as arguments reinforcing native speakers' intuitions and are not in a position to assess various analyses that have been proposed in the literature as alternatives to the classic analysis.

Parsed Corpus for Testing Generalizations in Japanese Syntax / 5  $\,$ 

NQ floating is possible with nominative and accusative arguments, but not with arguments marked with postpositions such as *kara* 'from', *de* 'with', etc. The example in (2a) is a case where a numeral quantifier *san-nin* 'three (persons)' appears inside the host nominative subject, and (2b) represents a case where it is placed to the immediate right of the host argument by NQ floating.

- (2) a. **[San-nin-no** sensei-ga] osake-o non-da. three-CL-GEN teacher-NOM sake-ACC drink-PAST 'Three teachers drank sake.'
  - b. Sensei-ga **san-nin** osake-o non-da. teacher-NOM three-CL sake-ACC drink-PAST 'Three teachers drank sake.'

The same NQ floating operation may apply to an accusative argument. (3a) is a case where the numeral quantifier *san-bon* 'three (bottles)' occurs inside the accusative argument, but in (3b), it appears after the accusative argument, i.e. outside its associated host argument by NQ floating.

(3)	a.	Sensei-ga	[san-bon-no	osake-o]	non-da.	
		teacher-NOM	${\rm three\text{-}CL\text{-}GEN}$	${\it sake-ACC}$	$\operatorname{drink-PAST}$	
		'The teacher drank three bottles of sake.'				

b. Sensei-ga osake-o **san-bon** non-da. teacher-NOM sake-ACC three-CL drink-PAST 'The teacher drank three bottles of sake.'

In (2a) and (3a), the numeral quantifiers are marked with genitive case, indicating that they are included in the host nominals. On the other hand, in (2b) and (3b), the numeral quantifiers appear to the right of the host nominals without genitive marking, showing that they are floated off their hosts.<sup>5</sup>

- (i) a. Sensei-ga <u>kinoo</u> san-nin osake-o non-da. teacher-NOM yesterday three-CL sake-ACC drink-PAST 'Three teachers drank sake yesterday.'
  - b. Sensei-ga osake-o <u>kinoo</u> san-bon non-da. teacher-NOM sake-ACC yesterday three-CL drink-PAST 'The teacher drank three bottles of sake yesterday.'

 $<sup>{}^{5}</sup>$ The fact that the floated numeral quantifiers are located outside the host arguments can be ascertained that an adverb like *kinoo* 'yesterday' can intervene between the two, as shown in (i).

6 / LiLT volume 18, issue 2

Floated numeral quantifiers do not necessarily have to appear in a position contiguous with their hosts. When numeral quantifiers are placed before the verb, there arises an asymmetry in acceptability, as shown in (4).

- (4) a. Osake-o sensei-ga **san-bon** non-da. sake-ACC teacher-NOM three-CL drink-PAST 'The teacher drank three bottles of sake.'
  - b. \*Sensei-ga osake-o **san-nin** non-da. teacher-NOM sake-ACC three-CL drink-PAST 'Three teachers drank sake.'

In (4a), the object appears in sentence-initial position. The numeral quantifier san-bon occurs in a position separate from the object, but the sentence is legitimate. In (4b), by contrast, the numeral quantifier san-nin placed in front of the verb cannot be taken to be floated off the subject.

The data show that a numeral quantifier can be placed before the verb if it is launched off the object, but not the subject. As discussed by Miyagawa (1989), the fact of quantifier floating in (4a) can be accounted for if the object first appears before the verb and is moved to the front of the sentence.<sup>6</sup>

- (ii) a. Sensei-ga osake-o koremade-ni san-nin non-da. teacher-NOM sake-ACC since three-CL drink-PAST
   'Three teachers drank sake since then.'
  - b. Gakusei-ga sono tan'i-ga (seikaku-ni) 49-nin hosi-i.
     student-NOM that credit-NOM precisely 49-CL want-pres
     'Precisely 49 students want that credit.'

There are alternative analyses such as Gunji and Hasida (1999) for the phenomena, but the judgments are apparently subject to a wide range of speaker variation for some examples. In addition, since no general agreement is reached on alternative analyses, we will stick to the argument based on Kuroda's observation in the paper.

<sup>6</sup>To be more exact, Miyagawa (1989) argues that floated numeral quantifiers can be associated with their host as long as they have a mutual c-commanding relation.

These examples illustrate that the floated numeral quantifiers are not included inside their associated host arguments. One reviewer raises the question of some alternative accounts are needed in the presence of examples like (i), which apparently run counter to the baseline generalization.

PARSED CORPUS FOR TESTING GENERALIZATIONS IN JAPANESE SYNTAX / 7

(5) a. [ SUBJ [AOBJ Bsan-bon V]T] b.  $[OBJ_{I} SUBJ [COBJ_{I} Dsan-bon V]T]$ c. \*[ ESUBJ [OBJ Fsan-nin V]T]

If 4a, the numeral quantifier *san-bon* can appear in a position separated superficially from the scrambled object. The numeral quantifier can be placed in the pre-verbal position since a copy of the object left by movement appears in object position, as represented in (5b). On the other hand, the numeral quantifier *san-nin* cannot be associated with the subject in (4b), since no copy of the subject appears in the preverbal position with no movement involved, as represented in (5c). The contrast in acceptability between (4a) and (4b) suggests then that the nominative-accusative order is basic, and the accusative-nominative order is derived.<sup>7</sup>

Another argument on the ordering of arguments may be derived from facts regarding quantifier scope. As often observed (e.g. Kuroda 1970),

(i) Otyuugen-wa [NP sake-ga kyonen 3-bon] to [biiru-ga mid.summer.gift-TOP sake-NOM last.year 3-CL and beer-NOM kotosi 12-hon] dat-ta. this.year 12-CL COP-PAST
'The mid-summer gifts were three bottles of sake last year and 12 bottles of beer this year.'

Aside from slightly deteriorated acceptability of the example (at least for some speakers), we should note that Fukui and Sakai (2003) argue that this kind of data do not present arguments for Fukushima's claim that to conjoins nominals exclusively. Fukushima's argument is suspect even on an empirical level, since adverbs included in the alleged nominals are not licensed unless some verbal projections are present in them. The same reviewer also asks whether the facts of transitive clauses obtain under the VP-internal subject assumption. Although we do not discuss exactly how they are derived, we can note that if the split-vP analysis, where VP is divided into two verbal projections, is adopted, the facts can be accounted for under the notion of c-command, even with this particular theoretical assumption. This is because we have the representation: [SUBJ<sub>i</sub> OBJ<sub>j</sub> [t<sub>i</sub> [san-nin t<sub>j</sub> V] v] T] rather than [SUBJ<sub>i</sub> OBJ<sub>j</sub> [t<sub>i</sub> san-nin t<sub>j</sub> V] T].

<sup>&</sup>lt;sup>7</sup>Some complications arise since, as argued by Miyagawa (1989), floating quantifiers can be attached to the trace created by A-movement. This means that different results arise if we consider clauses in which D- and S-structures do not match. Since we are only concerned with the 'shallow' ordering generalizations, we do not discuss cases where D-structure representations are derived by A-movement, such as unaccusative and passive clauses, where subjects are claimed to originate from object positions. One reviewer remarks that an example like (i), citing Fukushima (2003), show that a floating quantifier is not necessarily independent from its host.

**JULY 2019** 

a sentence like (6a) is unambiguous, but a scrambled sentence like (6b) is ambiguous with regard to the relative scope of the two quantifiers.

- (6) a. Dareka-ga daremo-o aisi-te i-ru. someone-NOM everyone-ACC love-GER be-PRES 'Someone loves everyone.' (some>every)
  - b. Daremo-o dareka-ga aisi-te i-ru. everyone-ACC someone-NOM love-GER be-PRES 'Someone loves everyone.' (some>every, every>some)

In the transitive clause in (6a) with a nominative subject and an accusative object, which represents the non-derived order of arguments, the only scope interpretation available is one in which the existential quantifier takes scope over the universal quantifier.<sup>8</sup> On the other hand, the scrambled version of (6a), which is given in (6b), has scope ambiguity with either of the two quantifiers taking scope over the other.

It is important to see that in Japanese, quantifier scope observes the so-called 'rigidity' condition. This condition states that the order of quantifier arguments is reflected in their scope interpretation (Hoji 1985). This amounts to declaring that when no scrambling is involved, the sentence has only the scope interpretation that reflects the surface order of arguments. Under this perspective, the differences in scope interpretation between (6a) and (6b) may be accounted for by positing the structures in (7).

(7) a. [TP dareka-ga [VP daremo-o V] T]
b. [TP daremo-o [TP dareka-ga [VP daremo-o V] T]]

For expository purposes, we assume that scrambling is an adjunction operation (Saito 1985), and thus, (6b) is derived from (6a) by moving the accusative object *daremo* 'everyone' across the nominative subject *dareka* 'someone' and adjoining it to TP, as illustrated in (7b). We can postulate here that the scope of quantifiers is fixed with reference to (overt or covert) copies of quantifiers available at the surface level.

<sup>&</sup>lt;sup>8</sup>One reviewer reports that different scope interpretations are obtained when different quantifiers, e.g. *dono* 'which', are chosen. It is a general observation that scope interpretations vary depending on the type of quantifier, and it is known that quantifiers of the *dono*-type show peculiar behaviors (see e.g. Pesetsky 1987). Thus, for illustrative purposes, we are using more standard quantifiers i.e. *dare* 'who' and *nani* 'what' types of quantifiers, rather than *dono*-quantifiers. Even these quantifiers, speaker variation is found. For these reasons, we will follow the standard judgments to discuss quantifier scope.

Parsed Corpus for Testing Generalizations in Japanese Syntax / 9  $\,$ 

When no scrambling takes place, as in (6a), the subject quantifier dareka 'someone' is taken to be located in a higher position than the object quantifier daremo 'everyone', as represented in (7a). Thus, the only interpretation available for (6a) is that dareka takes scope over daremo (some>every). If the object is scrambled across the subject, as in (6b), the structure given in (7b) is obtained. If the overtly realized object daremo is taken to determine its scope, daremo takes scope over dareka (every>some) On the other hand, if the covert copy in the object position is taken to be relevant for its scope determination, the quantifier dareka takes scope over daremo (some>every). Thus, when scrambling is involved, the sentence turns out to be ambiguous. This fact suggests that (6a) represents the basic order, and (6b), a derived one.

The arguments reviewed above illustrate that in canonical transitive clauses, the basic word order is the one in which nominative subjects precede accusative objects, and not the other way around. With this generalization in mind, let us now turn to the question of what we can say about the ordering of arguments with the help of its usage frequency in the corpora.

Some remarks on corpus in order at this point. For the purpose of adducing the samples of various types of clauses, we searched Keyaki Treebank (Butler et al. 2012, 2018) consisting of 58842 sentences (accessed 3/14/2018 through 4/14/2018). The targets of the corpus search are simple transitive clauses, and complex clauses such as causative and passive clauses are excluded to avoid possible complexities in analyzing the data. Furthermore, the corpus search is restricted to clauses which include case-marked subjects and objects in them, and clauses containing subjects or objects without overt case-marking (such as 'X-wa ... Y-o' and 'X-ga ... Y-mo') are not counted for their grammatical relations are not overtly marked. This ensures that the collected data represent the case-marking patterns of various types of basic clauses (with or without scrambling).<sup>9</sup>

Cases involving canonical transitive clauses with nominative subjects and accusative objects are straightforward. The figures in (8) represent search results of clauses with nominative subjects and accusative ob-

<sup>&</sup>lt;sup>9</sup>Token counts show the tendency of argument ordering in representative basic clauses in naturally-occurring data. They are not designated to pinpoint possible factors affecting the tendency. Some of such possible factors are discussed by checking the actual examples extracted from the corpus. One reviewer asks whether there is any difference between different genres. To our knowledge, there is no such difference, and genre is not an issue as far as scrambling is concerned.

jects using Tregex (https://nlp.stanford.edu/software/tregex.shtml).<sup>10</sup>

(8) <u>Canonical Transitive Construction</u>
 A: Nom.Subject-Acc.Object: 8269 (98.8%)
 B: Acc.Object-Nom.Subject: 104 (1.2%)

It is often claimed that scrambling can apply rather freely. In fact, since, in most sentences, an object can be scrambled across the subject without affecting grammaticality, this claim sounds reasonable. Nevertheless, sentences with scrambled objects are far less common than sentences without scrambling in naturally-occurring examples in the corpus; overwhelmingly, subjects occur before objects, as shown in (8).

This fact would be naturally expected if the basic word order of a canonical transitive clause is 'nominative.subject-accusative.object'. Moreover, there is also good reason to think that this 'nominative.subjectaccusative.object' order is basic in terms of information flow. Broadly speaking, subjects refer to what is talked about and tend to have old information (or shared information) in discourse status. The ordering of arguments is in accordance with information structure, i.e. the flow of information, so the reversed order of object-subject does not often occur.

In our corpus samples, when the accusative-nominative order is obtained, long objects tend to be placed before short subjects, although this is not an absolute rule (see Yamashita and Chang 2001). In fact, our sample data indicate that there is no significant difference in the length of arguments when the nominative.subject-accusative.object order is obtained, while in 60% of the examples with scrambled accusativenominative order, objects are longer than subjects, and about 40% of the sentences with the reversed order include objects modified by noun-modifying clauses. Given this, it can be hypothesized that the scrambled order is often motivated by processing reasons (e.g. to avoid center embedding, which causes difficulty in sentence processing). In any case, what we can say with the help of empirical corpus counts in (8) is that the basic nominative-accusative order is overwhelmingly dominant.

#### 3.2 Non-canonical transitive clauses

Let us now turn to the non-canonical transitive clauses. In Japanese, while the nominative subject-accusative object pattern is canonical,

 $<sup>^{10}{\</sup>rm This}$  can be done by searching examples where IP, which signifies a clause, dominates both nominative and accusative arguments.

Parsed Corpus for Testing Generalizations in Japanese Syntax / 11  $\,$ 

some transitive predicates take a dative subject and a nominative object (complement), and others take a nominative subject and dative object (complement), as summarized in Table 1.

SUBJECT	OBJECT (COMPLEMENT)
DATIVE	NOMINATIVE
NOMINATIVE	DATIVE
NOMINATIVE	NOMINATIVE

TABLE 1: Non-canonical (mono)transitive clauses

When dealing with the non-canonical case-marking cases, it is necessary to confirm the grammatical relations of arguments, because dative arguments could be either subjects or complements, and nominative arguments could be either subjects or complements. Scrambling may apply to non-canonical transitive clauses, except when the arguments are both marked with nominative case.

# **3.2.1.** Dative-subject construction (dative.subject-nominative.object alignment)

A first case involving non-canonical case-marking patterns concerns dative-subject constructions. The example in (9) represents a case of dative subject construction, which has a dative subject and a nominative object.

(9) Ken-ni kodomo-ga sikar-e-ru. Ken-DAT child-NOM scold-POTEN-PRES 'Ken can scold children.'

Dative-subject predicates take experiencer subjects or possessor subjects, depending on the meaning the predicates express. Dative-subject predicates can be verbal (e.g. *dekiru* 'can do') or adjectival (e.g. *hi-tuyooda* 'necessary'); some of them are perception predicates (*mieru* 'see'), and others are potential predicates (*kakeru* 'can write'). In addition, transitive predicates taking dative subjects are mostly stative (Kuno 1973).<sup>11</sup>

In Japanese transitive clauses, when the subject is marked with dative case, its object must be marked with nominative case rather than accusative case, as (10) illustrates.

 $<sup>^{11}\</sup>mathrm{A}$  small number of non-stative predicates such as umareru 'be born' take dative subjects. See Kishimoto (2016).

#### 12 / LiLT volume 18, issue 2

(10) Ken-ni kodomo-ga/\*-o sikar-e-ru. Ken-DAT child-NOM/-ACC scold-POTEN-PRES 'Ken can scold children.'

With many transitive predicates taking dative subjects, subjects can have nominative case marking as an alternative case-marking option, but in this section, we are only concerned with cases where subjects are marked with dative case.

In non-canonical case-marking constructions, overt case marking is not a reliable way of diagnosing the grammatical relations of arguments. Accordingly, we need to resort to some diagnostic tests to confirm their grammatical status. In Japanese, two major subject diagnostics are available—subject honorification and reflexivization. These diagnostic tests provide evidence that in the dative-subject construction, the dative argument serves as a subject.

It is well-known that Japanese reflexive *zibun* 'self' has subject orientation, i.e. its antecedent is confined to arguments serving as subjects, as exemplified by a canonical transitive clause in (11).

(11) Ken<sub>i</sub>-ga zibun<sub>i</sub>-no heya-de kodomo-o sikat-ta. Ken-NOM self-GEN room-in child-ACC scold-PAST 'Ken scolded his child in his own room.'

In the canonical transitive clause in (11), the nominative argument can be the antecedent of *zibun*, but the accusative argument cannot, showing that the nominative argument is the subject. By contrast, in the dative-subject construction in (12), the dative argument can be the antecedent of reflexive *zibun* 'self', but not the nominative argument.

(12) Ken<sub>i</sub>-ni zibun<sub>i</sub>-no heya-de kodomo-ga sikar-e-ru. Ken-DAT self-GEN room-in child-NOM scold-POTEN-PRES 'Ken can scold his child in his own room.'

The example illustrates that the dative argument serves as a subject. The nominative argument cannot be the antecedent of *zibun*, since it does not have grammatical status as a subject.

Subject honorification is another well-known diagnostic; only subjects are legitimate targets for subject honorification. Thus, in a canonical case-marking transitive clause, subject honorification is targeted at its nominative argument, but not an accusative argument, as shown in (13). Parsed Corpus for Testing Generalizations in Japanese Syntax / 13  $\,$ 

(13) a. Kimura-sensei-ga ano hito-o Kimura-teacher-NOM that man-ACC o-home-ni-nat-ta. HON-praise-DAT-become-PAST 'Professor Kimura praised that man.'
b. \*Ano hito-ga Kimura-sensei-o that man-NOM Kimura-teacher-ACC o-home-ni-nat-ta.

HON-praise-DAT-become-PAST

'That man praised Professor Kimura.'

In the dative subject construction, the dative-marked argument, but not the nominative-marked argument, can be a target of subject honorification, as shown in (14).

(14)	a.	Kimura-sensei-ni			ano	hito-ga	mie-te	
		Kimu	ra-teacher-I	DAT	that	man-NOM	see-GER	
		irassayar-u. be.HON-PRES						
		'Professor Kimura can see that man.'						
			hito-ni			0	mie-te	

b. And into-in Kimura-sensei-ga inte-te that man-DAT Kimura-teacher-NOM see-GER irassayar-u. be.HON-PRES

'That man can see Professor Kimura.'

The examples illustrate that the dative argument, but not the nominative argument, serves as the subject in the dative-subject construction.

In dative-subject constructions, the dative subject, in unmarked cases, precedes the nominative object. Besides, (15) shows that a numeral quantifier can appear after the dative subject even when the host nominative object precedes the dative subject.

- (15) a. Ken-ni hon-ga 2-satu yom-e-ta. Ken-DAT book-NOM 2-CL read-POTEN-PAST 'Ken was able to read two books.'
  - b. Hon-ga Ken-ni 2-satu yom-e-ta. book-NOM Ken-DAT 2-CL read-POTEN-PAST 'Ken was able to read two books.'

Since the numeral quantifier 2-satu can appear in the preverbal position in (15b), just like (15a), it must be the case that the nominative object has been scrambled to the sentence-initial position in (15b). The fact suggests that the unmarked word order is 'dative.subject-nominative.object', and the reverse order is derived.

Let us now turn to token counts. (16) illustrates the token frequency counts of the 'dative-nominative' and the 'nominative-dative' case-marking patterns of the dative subject construction, together with a list of predicates found in the searched corpus examples.<sup>12</sup>

(16) Dative-Subject Construction

A: Dative SUBJ – Nominative OBJ: 56 (86.2%)
B: Nominative OBJ – Dative SUBJ: 9 (13.8%)
Predicates: aru 'be', iru 'be', hituyooda 'necessary', dekiru 'can do', wakaru 'understand', kakeru 'lack', nai 'not be', mieru 'see', kakasenai 'essential'

The figures show that the subject tends to precede the object, although the number of examples available in the parsed corpus is limited. The dative-nominative pattern is the unmarked word order, hence we have more tokens of the dative-nominative order. The percentage of the marked 'nominative.object-dative.subject' order in the dative subject constructions is significantly higher than the percentage of the marked 'accusative.object-nominative.subject' order in the canonical transitive constructions. Nevertheless, the order of dative-nominative dative pattern, suggesting that the dative-nominative word order is basic. In our corpus data, the reversed 'nominative.object-dative.subject' order is obtained when we have long objects, i.e. in all the examples with the reversed nominative.object-dative.subject order, nominative objects are longer than dative subjects.<sup>13</sup>

<sup>13</sup>In eight examples out of nine, the dative subject is a pronoun.

<sup>&</sup>lt;sup>12</sup>Potential verbs with the potential suffix can be formed productively, but no such verbs showing the dative-nominative pattern are found in the corpus. This is a coincidental fact since potential verbs with other case-marking patterns are also not commonly found (perhaps, due to the size limitation of the corpus). It should also be noted that dative arguments are awkward without wa (see Kishimoto et al. 2015); many instances of dative arguments include wa, since dative marking usually survives even with wa attached to the arguments, unlike nominative and accusative arguments. This is also true of the extracted examples of dative constructions. Our corpus data in fact include many dative-subject verbs taking wa-marked dative arguments. This wa marking does not affect the possibility of scrambling, for many instances of dative arguments include wa marking, regardless of whether they precede or follow the nominative objects.

Parsed Corpus for Testing Generalizations in Japanese Syntax / 15  $\,$ 

# **3.2.2.** Dative complement constructions (nominative.subject-dative.complement alignment)

There are also predicates (like au 'meet') taking 'nominative-dative' case-marking patterns. This type of construction is considered to be a variant of the nominative-subject construction, where the complement is marked with dative case rather than accusative case. Two representative examples are given in (17).

- (17) a. Ken-ga Mari-ni at-ta. Ken-NOM Mari-DAT meet-PAST 'Ken met Mari.'
  - b. Ken-ga Mari-ni yasasi-i. Ken-NOM Mari-DAT kind-PRES 'Ken is kind to Mari.'

Predicates taking nominative-dative patterns are mostly non-stative predicates (e.g. *au* 'meet', *ataru* 'hit'), but some are stative (e.g. *yasasii* 'kind', *kibisii* 'strict'). Note that the complements of transitive adjectives are marked with dative case, i.e. transitive adjectives do not take accusative complements, possibly with the exception of desideratives like *hosii* 'want', although there is speaker variation as to the admissibility of an accusative complement (see Shibatani 1978).

Let us now look at how reflexive zibun behaves in this type of construction. The examples in (18) show that the nominative argument can be the antecedent of reflexive zibun regardless of whether the predicate is non-stative or stative.

- (18) a. Ken<sub>i</sub>-ga Mari-ni zibun<sub>i</sub>-no heya-de at-ta. Ken-NOM Mari-DAT self-GEN room-in meet-PAST 'Ken met Mari in his own room.'
  - b. Keni-ga zibuni-no kodomo-ni yasasi-i. Ken-NOM self-GEN child-DAT kind-PRES 'Ken is kind to his own child.'

The data show that both non-stative and stative constructions may take nominative subjects and dative complements.

Subject honorification patterns with reflexivization, in that only the nominative argument can be the target, as shown in (19).

- (19) a. Kimura-sensei-ga gakusei-ni Kimura-teacher-NOM student-DAT o-ai-ni-nat-ta. HON-meet-DAT-become-PAST 'Professor Kimura met the student.'
  - b. Kimura-sensei-ga gakusei-ni Kimura-teacher-NOM student-DAT
    o-yasasi-kat-ta.
    HON-kind-PAST
    'Professor Kimura was kind to the students.'

These facts remain invariant regardless of whether the predicate is nonstative or stative and show that the nominative arguments are subjects in the dative-complement constructions.

Turning now to the corpus counts of transitive constructions taking nominative subjects and dative complements, the search results (their token counts and a list of predicates found in the corpus) are shown in (20).

 (20) Dative-Complement Construction

 A: Nominative SUBJ-Dative OBJ: 10 (100%)
 B: Dative OBJ-Nominative SUBJ: 0 (0%)
 Predicates: au 'meet', hairu 'enter', iu 'say'<sup>14</sup>, hanasikakeru 'talk'

In the corpus data (Keyaki treebank) we have accessed, no examples involving stative predicates are found. In addition, there are no examples with the nominative-dative order in this type of construction. Since the number of tokens is fairly small, these facts may be a coincidence, but the token counts in (20) illustrates, at least, that the nominative-dative order can be regarded as basic.

# **3.2.3.** Double nominative constructions (nominative.subject- nominative.complement alignment)

Besides the dative-subject predicates, there are predicates that take nominative-nominative case-marking patterns. Stative predicates including potential predicates often take a 'nominative-nominative' pattern, alongside a 'dative-nominative' pattern. (21) provides one representative example.

 $<sup>^{14}{\</sup>rm The}$  attested example of iu 'say' with this case frame has a to-complement, in addition to a dative complement.

Parsed Corpus for Testing Generalizations in Japanese Syntax / 17  $\,$ 

(21) Ken-ga heya-ga hosi-i. Ken-NOM room-NOM want-PRES 'Ken wants a room.'

Some transitive predicates like *sukida* 'like' take only the nominativenominative case-marking pattern.<sup>15</sup> Many dative subject predicates are allowed to take this case-marking pattern as well. Shibatani (1978) reports that there is speaker variation as to which predicate allows this case-marking pattern.

It is worth noting here that in double nominative constructions, scrambling the second nominative argument across the first nominative argument to the sentence initial position results in an ungrammatical sentence, as shown in (22).

(22) \*Heya-ga Ken-ga hosi-i. room-NOM Ken-NOM want-PRES 'A room, Ken wants.'

Japanese has a grammatical restriction that a nominative argument cannot be scrambled across another nominative argument.<sup>16</sup> Consequently, the scrambled sentence in (22) is not grammatical, even though the meaning is fully transparent with no possibility of confusion in semantic terms.

In double nominative constructions, the first nominative argument serves as the subject. Again, this can be confirmed by reflexivization. (23) shows that the first argument can be the antecedent of reflexive *zibun*.

(23) Ken<sub>i</sub>-ga zibun<sub>i</sub>-no heya-ga hosi-i. Ken-NOM self-GEN room-NOM want-PRES 'Ken want his own room.'

The example in (24) with subject honorification also illustrates that the first nominative argument can be a target of honorification.

<sup>&</sup>lt;sup>15</sup>The surface 'nominative-nominative' case-marking pattern may be derived from an intransitive predicate by applying a grammatical operation, which is sometimes referred to as 'subjectivization' (Kuno 1973). These cases are excluded from our discussion. We are only concerned with transitive clauses in this paper.

 $<sup>^{16}</sup>$ The constraint pertains to scrambling. When a nominative object is marked with wa, it can be moved across the nominative subject.

(24) Kimura-sensei-wa sore-ga totemo hosiku-te Kimura-teacher-TOP it-NOM very.much want-GER irassyar-u.
be.HON-PRES
'Professor Kimura wants it very much.'

In the double nominative construction, the first nominative argument serves as the antecedent of *zibun* and can be targeted for subject honorification, suggesting that the first nominative argument serves as the subject of this construction.

Let us now turn to the question of how the two nominative arguments of the double nominative construction are arranged. (25) shows the token counts of this construction, as well as its predicates, found in the corpus.

ʻlike'
4

The zero count of the 'nominative.object-nominative.subject' order comes as no surprise, given the restriction that no scrambling is allowed for the type of clause taking nominative-nominative case marking (see (22)).

### 3.3 Ditransitive construction

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 another. Such ditransitive predicates select dative-marked goal and accusative-marked theme arguments, alongside nominative subjects.

(26)	a.	John-ga	Mary-ni	hon-o	age-ta.		
		John-NOM	Mary-DAT	$\operatorname{book-ACC}$	give-PAST		
		'John gave	Mary a book.'				

b. John-ga Mary-ni hon-o okut-ta. John-NOM Mary-DAT book-ACC send-PAST 'John sent Mary a book.'

With regard to the word order of the internal arguments, Hoji (1985) claims that the 'dative-accusative' order is basic, but Miyagawa (1997)

claims that either the 'dative-accusative' or 'accusative-dative' order could be basic. (No claim has been advanced in the literature that the accusative-dative order is basic, while the dative-accusative order is derived.) Native speakers are likely to judge the dative-accusative order as basic, but their judgments are not as secure as transitive cases.

Let us now turn to arguments available in the literature that are claimed to determine the order of the internal arguments of ditransitive verbs. First, Hoji (1985) suggests that the indirect object is located above the direct object, on the basis of examples like those given in (27).

- (27) a. \*Kimi-wa [[ $e_i \ e_j$  F-o tuke-ta] sensei<sub>i</sub>]-ni dare<sub>j</sub>-o you-TOP F-ACC give-PAST teacher-DAT who-ACC aw-ase-ta no? meet-CAUS-PAST Q 'Who<sub>j</sub> did you have meet the teacher who gave him<sub>j</sub> an F?'
  - b. Kimi-wa dare<sub>i</sub>-ni [[e<sub>i</sub> e<sub>j</sub> okut-ta] ningyoo<sub>j</sub>]-o you-TOP who-DAT send-PAST doll-ACC kaesi-ta no? return-PAST Q
    'Who<sub>i</sub> did you return the doll that he<sub>i</sub> sent over to you to?'

The grammatical judgments in (27) are based on whether an unrealized null pronoun e can have a bound variable interpretation. The (un)availability of a bound variable interpretation depends on where a zero pronoun is located relative to a quantifier (or a wh-phrase) in the clause, i.e. a bound-variable interpretation is available for a zero pronoun provided it is c-commanded by its antecedent quantifier/whphrase. Observe that this bound pronoun interpretation is available in (27b), where a bound pronoun is included in the accusative object and the wh-phrase is the dative object, but not in (27a), where the dative object contains a bound pronoun and the wh-phrase is the accusative object. In light of this fact, Hoji argues that the indirect object is located in a higher position than the direct object. In the present perspective, this means that ditransitive verbs have the dative-accusative order as their basic word order.

Miyagawa and Tsujioka (2004) argue for the view that both dativeaccusative and accusative-dative order could be basic. While they present a number of arguments, we concentrate in their arguments based on idioms here.

- (28) a. Taroo-wa kuruma-o te-ni ire-ta. Taro-TOP car-ACC hand-DAT put.in-PAST 'Taro acquired a car.'
  - b. Taroo-wa genkoo-ni te-o ire-ta. Taro-TOP draft-DAT hand-ACC put.in-PAST 'Taro revised the draft.'

The idiom *te-ni ireru* 'acquire' in (28a) has the dative argument placed after the accusative argument. On the other hand, the idiom *te-o ireru* 'revise' in (28b) has an accusative argument preceded by the dative argument. On the assumption that the frozen order of idioms reflects basic word order, Miyagawa and Tsujioka (2004) claim that both the dative-accusative and the accusative-dative orders could be basic.<sup>17</sup>

With the two claims advanced in the literature in mind, let us proceed to discuss the token counts we have in the corpus data. Our corpus counts shown in (29) indicate that ditransitive predicates allow either the 'dative-accusative' or 'accusative-dative' order, but that the former is more frequent than the latter.<sup>18</sup>

<sup>18</sup>Constructions with predicates like *suru* 'make', which express the meaning of transformation, are excluded from the counts, although they look like having two internal arguments; one argument marked with ni and the other with o.

 (i) Ken-ga kodomo-o isya-ni si-ta. Ken-NOM child-ACC doctor-DAT do-PAST 'Ken made his child a doctor.'

There are two reasons for this exclusion. For one thing, there is an issue over whether the ni-marking occurring with the nominal can be construed as a dative marker or not. Okutsu (1978) claims that ni is an instance of copula. If this is the case, nicannot be dative marker. For another, this verb does not allow its internal arguments to be permuted, as shown in (ii).

 (ii) \*Ken-ga isya-ni kodomo-o si-ta. Ken-NOM doctor-DAT child-ACC do-PAST 'Ken made his child a doctor.'

Many instances of the causative verb *suru* taking this pattern are found in the corpora. Nevertheless, this causative construction does not show word order variation, so the token counts of the causative construction are not included in the search counts in (29).

<sup>&</sup>lt;sup>17</sup>To be more precise, Miyagawa and Tsujioka (2004) claim that the ordering possibilities differ depending on whether the dative argument counts as a high goal or a low goal. Although we do not go into the details of their discussion, their main claim is that ditransitive verbs can have two different alignments of internal arguments via base-generation.

Parsed Corpus for Testing Generalizations in Japanese Syntax / 21

(29) <u>Ditransitive Construction</u>

A: Dative OBJ – Accusative OBJ: 209 (61.8%)

B: Accusative OBJ – Dative OBJ: 129 (38.2%)

Predicates: ageru 'give', watasu 'hand', ataeru 'give', hookoku-suru 'report'

The difference between the two is fairly small, compared with the transitive cases discussed earlier, with the approximate ratio of 6 to 4. 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. Given the theory of markedness, however, it can be stated that the dative-accusative order of internal arguments is basic in ditransitive constructions, since this dative-accusative ordering has more tokens that the other accusative-dative order (see also Koizumi and Tamaoka 2004 for an experimental study, which concludes that the dative-accusative order is basic).<sup>19</sup>

As discussed earlier, in mono-transitive clauses, longer objects tend to be placed before shorter subjects. In ditransitive clauses, a different factor plays a role in determining the order of internal arguments, because in our corpus data, the length of the internal arguments has no bearing on the ordering. Notably, the internal arguments of ditransitive constructions often include discourse-anaphoric pronouns (including demonstrative pronouns (e.g. *sore* 'it') and personal pronouns (*kare* 'he', *kanozyo* 'she')) regardless of the order of the internal arguments. Specifically, in ditransitive constructions with dative-accusative order, dative arguments including discourse-anaphoric pronouns account for 13.6%, and accusative arguments containing discourse-anaphoric pronouns, 1.1%. With the reversed accusative-dative order, accusative objects including discourse-anaphoric pronouns represent 36.3%, and dative arguments containing discourse-anaphoric pronouns, 1.3%.<sup>20</sup>

Since anaphoric pronouns are indicative of the discourse-old status of their associated arguments (see Birner and Ward (1998)), the data suggest that the internal arguments with discourse-old information status tend to precede the other internal arguments and in particular,

<sup>&</sup>lt;sup>19</sup>Sasano and Okumura (2017) has reached the same conclusion. As noted by one researcher, conflicting results are obtained in some experimental studies on ditransitive predicates, and Koizumi and Tamaoka is not the only view.

<sup>&</sup>lt;sup>20</sup>This discourse factor does not affect the ordering of arguments in canonical transitive clauses. This can be seen by the fact that in the first 100 examples with the nominative-accusative order in the corpus tokens, 3% of nominative subjects and 12% of accusative objects include anaphoric pronouns. In the 104 examples with the reversed accusative-nominative order, 1% of nominative subjects and 15% of accusative objects include anaphoric pronouns.

that reordering of dative and accusative arguments is motivated by the discourse-old status of the fronted accusative arguments. The facts of ditransitive predicates can be taken as a good indication that the two internal arguments of ditransitive verbs can be easily reordered when an accusative argument carries discourse-old information.

#### 4 Summary

In this case study, we found that in naturally-occurring corpus data, a significantly large number of occurrences of sentences with SOV word order are found, while the occurrences of sentences with OSV word order are much smaller in number. Although scrambling is possible in many clauses, the attested examples with scrambled order are restricted in number. Scrambled sentences account for a fraction of tokens, 1.2% in canonical transitive clauses and 15% in dative-subject constructions.

In ditransitive constructions, we found more examples with the dative-accusative order than those with the accusative-dative order in the naturally-occurring corpora. The percentage of the 'accusative-dative' order in ditransitive constructions is much higher than that of the derived 'accusative-nominative' order in canonical transitive constructions and the derived 'nominative-dative' order in dative-subject constructions. The higher percentage of the reversed order in ditransitive constructions reflects the uncertainty of the native speaker's judgments about the internal arguments of ditransitive predicates, but the data suggest that the dative-accusative order can be regarded as the basic word order.

#### References

- Birner, Betty J. and Gregory Ward. 1998. Information Status and Noncanonical Word Order in English. Amsterdam: John Benjamins.
- Butler, Alastair, Tomoko Hotta, Ruriko Otomo, Kei Yoshimoto, Zhen Zhou, and Hong Zhu. 2012. Keyaki treebank: phrase structure with functional information for Japanese. In *Proceedings of Text Annotation Workshop*, *National Institute of Informatics (NII)*.
- Butler, Alastair, Kei Yoshimoto, Shota Hiyama, Stephen Wright Horn, Iku Nagasaki, and Ai Kubota. 2018. The Keyaki Treebank Parsed Corpus, Version 1.1.
- Fukui, Naoki and Hiromu Sakai. 2003. The visibility guidance for functional categories: Verb raising and related issues. *Lingua* 113:321–375.
- Fukushima, Kazuhiko. 2003. V-raising and numeral classifiers in Japanese: Incompatible bedfellows. Journal of East Asian Linguistics 12:313–347.
- Gunji, Takao and Kôiti Hasida. 1999. Measurement and quantification. In T. Gunji and K. Hasida, eds., *Topics in Constraint-Based Grammar of Japanese*, 36–97. Dordrecht: Kluwer.

- Hoji, Hajime. 1985. Logical Form Constraints and Configurational Structures in Japanese. Ph.D. thesis, University of Washington, Seattle.
- Kishimoto, Hideki. 2016. Stative and existential/possessive predicates. In T. Kageyama and H. Kishimoto, eds., *Handbook of Japanese Lexicon and* Word Formation, 559–598. Berlin: De Gruyter Mouton.
- Kishimoto, Hideki, Taro Kageyama, and Kan Sasaki. 2015. Valency classes in Japanese. In B. Comrie and A. Malchukov, eds., Valency Classes in the World's Languages, 765–805. Berlin: De Gruyter Mouton.
- Koizumi, Masatoshi and Katsuo Tamaoka. 2004. Cognitive processing of Japanese sentences with ditransitive verbs. *Gengo Kenkyu* 125:173–190.
- Kuno, Susumu. 1973. The Structure of the Japanese Language. Cambridge, MA: MIT Press.
- Kuroda, S.-Y. 1970. Remarks on the notion of subject with reference to words like also, even, or only. part II. Annual Bulletin, Research Institute of Logopedics and Phoniatrics 4:127–152.
- Kuroda, S.-Y. 1983. What can Japanese say about government and binding. WCCFL 2:153–164.
- Miyagawa, Shigeru. 1989. Syntax and Semantics 22: Structure and Case Marking in Japanese. San Diego: Academic Press.
- Miyagawa, Shigeru. 1997. Against optional scrambling. *Linguistic Inquiry* 28:1–25.
- Miyagawa, Shigeru and Takae Tsujioka. 2004. Argument structure and ditransitive verbs in Japanese. *Journal of East Asian Linguistics* 13:1–38.
- Okutsu, Keiichiro. 1978. Boku-wa Unagi-da no Bunpoo. Tokyo: Kurosio Publishers.
- Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In E. J. Reuland and A. ter Meulen, eds., *The Representation of* (*In*)Definiteness, 98–129. Cambridge, MA: MIT Press.
- Saito, Mamoru. 1985. Some Asymmetries in Japanese and Their Theoretical Implications. Ph.D. thesis, MIT, Cambridge, MA.
- Sasano, Ryohei and Manabu Okumura. 2017. Corpus-based analysis of the canonical word order of Japanese double object constructions. Journal of Natural Language Processing 24(5):687–703. Shorter version in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016), pp.2236–2244.
- Shibatani, Masayoshi. 1978. Nihongo-no Bunseki. Tokyo: Taishukan.
- Tomic, Olga Miseska. 1989. *Markedness in Synchrony and Diachrony*. Berlin: Mouton de Gruyter.
- Yamashita, Hiroko and Franklin Chang. 2001. 'Long before short' production in a head-final language. Cognition 81:B45–B55.

Resources used:

- Alastair Butler, Kei Yoshimoto, Shota Hiyama, Stephen Wright Horn, Iku Nagasaki, and Ai Kubota. 2018. The Keyaki Treebank Parsed Corpus, Version 1.1 (http://www.compling.jp/keyaki/).
- National Institute for Japanese Language and Linguistics (2016). "NINJAL Parsed Corpus of Modern Japanese." (Version 1.0) (http://npcmj.ninjal. ac.jp/interfaces/).