The Influence of Givenness and Heaviness on OSV in Japanese

Satoshi Imamura

Japanese Studies, University of Oxford
41 Wellington Square
Oxford, OX1 2JF
United Kingdom

Abstract

Several studies contend that the main motivation for scrambling is heaviness. In particular, Yamashita (2002) maintains that scrambling has nothing to do with givenness and that heaviness is the primary factor for scrambling. However, her conclusions count on only 19 examples and she does not distinguish VP-internal scrambling from VP-external scrambling. Thus, it is conceivable that some types of scrambling rely on givenness. In order to see if this hypothesis is on the right track, I conducted a corpus analysis of OSV order in Japanese, largely based on the quantitative approach. Consequently, it has been revealed that both givenness and heaviness have a high explanatory power for the usage of OSV order. Furthermore, there was no correlation between givenness and heaviness, showing their independent influence on OSV order. Therefore, I conclude that both givenness and heaviness are sufficient to trigger OSV order and the phenomenon cannot be fully accounted for except with reference to both. Furthemore, based on the mapping between information structure and syntactic structure, I propose that VP-external scrambling is discourse-driven while VP-internal scrambling is not.

1. Introduction

A natural language may have many kinds of options for expressing the same proposition. In Japanese, for example, the meaning of a canonical transitive sentence SOV can be expressed by a scrambled sentence OSV in which the object appears before the subject. Why do languages have many options to convey the same proposition? One explanation is that these options allow speakers to choose the way information is transmitted. They differ not in what is said about the world, but in the way it is packaged (Chafe, 1976; Lambrecht, 1996; Vallduvi and Engdahl, 1996). In other words, their differences derive from information structure, i.e. how the meaning of a sentence is conveyed. Specifically, it has long been recognized since the work of the Prague School that speakers prefer to put given information before new information. However, this description begs the question because givenness itself is not a clear-cut concept. Therefore, it is necessary to define givenness in an objective way. In this paper, givenness is defined by a quantitative approach (Givón, 1983) and regarded as discourse-old information i.e. information mentioned in the preceding discourse. In other words, previously mentioned constituents are considered to be given information. Another explanation for variable ordering of arguments is based on heaviness. Hawkins (1994) observed that long constituents tend to be put in earlier positions than shorter ones in Japanese in order to facilitate the processing cost of heavy constituents. In this study, I am going to investigate the usages of OSV

order in terms of givenness and heaviness, mainly based on quantitative data from a Japanese corpus.

This paper is organized as follows. Section 2 surveys previous studies about Givōnian givenness and scrambling, where I will overview the basic concepts of referential distance, given-new ordering, and heaviness. Section 3 presents my corpus analysis of scrambling from the viewpoint of information structure and heaviness. Then, I will reveal that O tends to be discourse-old information in OSV. In addition, I will demonstrate that heaviness has an effect on OSV order, independent of givenness. Moreover, from the viewpoint of mapping between syntactic structure and information structure, I propose that givenness will have greater effects on VP-external scrambling than on VP-internal scrambling. In contrast, heaviness seems to have stronger effect on VP-internal scrambling than on VP-external scrambling. Section 4 is devoted to the conclusion and further studies.

2. Previous Studies

2.1. Givōnian Givenness

Givōn (1983) proposes as one quantitative approaches for calculating the topicality of referents. The metric of Referential Distance (RD) measures the gap between a referent in the current clause and its antecedent using clause boundaries as units. If there is no antecedent in the previous clauses, RD is assigned a value of 20 because without some limitation it would be infinite [1]. Hence, RD is expressed by some number of clauses from 1 to 20. What I should emphasize here is that RD is a quantitative value and has several measures to assess degrees of givenness. That is, it is possible to state that some referent is

---

[1] The limitation of RD is rather arbitrary. For example, Givón (1994) proposed that it should be 3 and Cooreman (1992) suggested that it should be 15 because there was no example with RD higher than 15. However, we observed sentences with RD higher than 16, so we followed the criteria of Givón (1983).

older than other referents. Let us illustrate this concept with (1).

(1) a. I met a man on the road to Philadelphia.
     b. He had no face.
     c. Suddenly, he said to me
     d. that I would die soon.
     e. Somehow I thought
     f. that he told the truth.

In order to measure the RD of *he* in (1f), you need to go back to (1c). Since there are three clause boundaries between *he* in (1f) and *he* in (1c), RD for *he* in (1f) is 3. Although the same referent is once mentioned in (1a) and (1b), this has nothing to do with the RD of *he* in (1f). This is because RD is the value of the distance between the target referent and its nearest antecedent.

In this study, I will rely on RD for the purpose of calculating the givenness of scrambled objects. RD is a well recognized measurement that is easily implementable and its employment renders the results of my analysis reproducible.

2.2. Scrambling

In Japanese, it has been said that O in OSV is moved from the VP-internal position toward the sentence initial position (Miyagawa 2010; Saito, 1985, 2009). This phenomenon is called scrambling. Note that scrambling does not change the propositional meaning. What is the motivation for scrambling? One explanation is based on givenness. Kuno (1978:54) observed that word order choice in Japanese depends on given-new ordering, which means that given information is mentioned early and new information later. Applying this principle to OSV sentences, native Japanese speakers are thought to prefer OSV just in those cases where the direct object is more given than its subject. Saeki (1960) observed that NPs with demonstratives precede other constituents in general. This tendency is true of OSV. In particular,

Ishii (2001) observed that when scrambled objects are modified by demonstrative *sono* 'that', the acceptability of those sentences increases, as can be seen in the difference in acceptability between (2a) and (2b). Taken together, these studies suggest that there is a correlation between scrambled object and givenness in Japanese OSV word order.

(2) a. *okane-o   dare-ga   nusun-da-no?
        money-ACC   who-NOM   steal-PAST-Q
     b. sono okane-o   dare-ga
        that money-ACC   who-NOM
        nusun-da-no?
        steal-PAST-Q
        'Who stole that money?'

                                          (Ishii 2001: 97)

Another motivation for scrambling is the heaviness of the NP that is moved to the left. Yamashita and Chang (2001) revealed that native Japanese speakers were apt to shift long constituents to earlier positions more than short constituents in sentence production. This result is consistent with Saeki (1960), who observed that long NPs tend to precede short NPs. According to Hawkins (1994), the motivation for word order change is to facilitate the processing cost of heavy constituents.

Yamashita (2002) even insists that heaviness is more important for scrambling than referentiality is. In her written Japanese data, heaviness accounts for about 70% of the scrambled sentences while referentiality makes up about 25%. In other words, 70% of scrambled objects are long and 25% of them include a determiner or an anaphor either referring to something appearing in the preceding discourse, or inferable from it. She observed a complementary distribution between heaviness and referentiality because almost all referential constituents were light. However, her data include various types of scrambling: VP-internal, short-distance, and long-distance scrambling. Therefore, pure data are needed to examine the function of OSV in Japanese. Moreover, though Yamashita (2002) contends that heaviness is independent of referentially, it is not clear whether scrambled

heavy direct objects are discourse-old information or not. The range of discourse-old information is wider than referentiality because referential NPs must have a demonstrative or anaphor such as *sono* 'that' and *sonna* 'such' but there is no such constraint for discourse-old information. Therefore, it is conceivable that scrambled direct objects in OSV are both heavy and discourse-old. If one factor strongly depends on the other, that concept is not necessary for explaining the function of OSV order. In contrast, it is possible that givenness is unrelated to heaviness. This means that both concepts are needed to explain the function of scrambling. In this study, I am going to examine whether there is an interaction between heaviness and givenness in Japanese OSV word order.

To sum up the above discussion, there are two research questions that I attempt to solve in this study. The first question is whether O in OSV is given information or not. The second question is whether both givenness and heaviness independently affect OSV word order in Japanese, or both factors work together. On the basis of Givōnian approach, I will disentangle these issues.

3. Corpus Analysis of Scrambling

3.1. Basic Predictions and Procedure

The first aim of my study is to investigate the relationship between discourse-old information and OSV word order in Japanese. In order to attain my goal, I am going to calculate the RDs of objects in OSV. If the discourse status of direct object determines whether the speaker should use OSV or not, OSV is preferred when the RD of the direct object is less than 20. The second purpose of this study is to see if there is a correlation between givenness and heaviness. If there is a strong correlation between givenness and heaviness, one factor may be derived from the other. In contrast, if there is no correlation between them, this will mean that both concepts have an influence on OSV word order independently, showing autonomy of each concept. In order to check which hypothesis is more valid, I will measure the length of scrambled objects and compare it with their RDs.

3.2. Method

3.2.1. Corpus Data

The Balanced Corpus of Contemporary Written Japanese (BCCWJ) was employed in order to assemble relevant data. BCCWJ is designed to be representative of contemporary written Japanese and thus includes 100 million words from well-balanced written materials covering books, magazines, newspapers, library books, bulletin boards, blogs, best-selling books, school textbooks, minutes of the National Diet, publicity of newsletters of local governments, laws, and poetry verses (see Maekawa et al. 2008 in detail).

3.2.2. Materials

OSV sentences were collected from BCCWJ by using *Chunagon*, which is a web interface program. In particular, the string [*o*(ACC)-noun-*ga*(NOM)] was used to extract OSV examples. The reason for using only strings with subject NPs of minimal length is that the left boundaries of NPs are not marked in the corpus. The limitation of my design is that it cannot pick out complex subjects completely. Complex subjects modified by a relative in OSV like [[noun-*ga*-verb]-noun-*ga*] were eliminated by hand in order to control the data. Thus, the scope of the OSV string includes only a simple (non-branching) noun subject.

3.2.3. Calculation of Heaviness

In order to measure the lengths of direct objects, I counted the *bunsetsu* of direct objects. *Bunsetsu* is a basic linguistic unit in Japanese Linguistics, consisting of content word(s) followed by zero or more functional words. Generally speaking, *bunsetsu* corresponds to a phrase. The reason why I chose *bunsetsu* is that the length of the subjects in my study is controlled in terms of *bunsetsu*. The *bunsetsu* of the subjects is always 1 in my data because they are a single noun plus nominative case particle *GA*. In (3), for instance, *kuruma-ga* 'car-NOM' forms a *bunsetsu* because it is a content word *kuruma* 'car' followed by a functional word *GA*. As a whole, (3) consists of four *bunsetsu*s: *sono*, *kasao*, *kurumaga*, and *hanetobashita*.

(3) Sono kasa-o   kuruma-ga hanetobashi-ta.
    that umbrella-ACC car-NOM   hit-PAST
    'A car hit that umbrella.'

(BCCWJ)

3.2.4. Criterion of Given and New Information

In this study, a value along the scale of given-new is assigned according to the measurement of RD. When a constituent has its RD less than 20, it is regarded as discourse-old information. In contrast, when a constituent does not have an antecedent, it belongs to new information.

As for givenness, some kinds of inferable information are categorized into discourse-old information. In particular, bridging relations are taken into consideration. Bridging is an inference from a referent explicitly mentioned in the preceding discourse. In (4), the hearer must suppose that *ringo* 'apple' is a part of *kudamono* 'fruit'. This relation is a bridging relation. Though *ringo* 'apple' is not directly referred to in (4a), its RD is 1 because *kudamono* 'fruit' can be regarded as the antecedent.

(4) a. Taro-wa   kudamono-o kat-ta
        Taro-TOP fruit-ACC   buy-PAST
        'Taro bought fruit.'
    b. Shikashi, ringo-wa   kusattei-ta.
        but   apple-TOP   be.rotten-PAST
        'But the apples were rotten'

Yet, those examples which have no direct relationship with the previous discourse are not considered to be discourse-old information. In (5), both *football* and *baseball* belong to *sports*. Thus, *baseball* is indirectly connected with *football* through the concept *sports*. However, there is no direct relationship because *baseball* is not included in *football*. Therefore, baseball is not regarded as discourse-old information.

(5) a. Do you watch football?
     b. Yeah. Baseball I like a lot BETTER.
                    (Ward and Birner 1998: 161)

3.2.5. Criterion of RD Analysis

The criterions of my analysis are mainly based on Shimojō (2005), but several modifications are added to my analysis. In the following sections, I will explain the details of these criterions.

3.2.5.1. Complex Clause
Complex clauses are divided into separate clauses based on predicates. Therefore, subordinate clauses are regarded as independent clauses. For example, the complex clause (6) is divided into three clauses because it contains the three predicates; *kumu* 'pull up', *hayaokisusu* 'get up early', and *iu* 'say'.

(6) [$_3$ shin-iemoto-wa   musuko-kara
       new.head.of.school-TOP   son-from
    [$_2$ ojiichan-ni   sakini
       grandfather-DAT   in.first
       kuma-re-nai-youni]   [$_1$(S)
       pull.up.PASS-NEG-so.as.to   (he)
       hayaoki-shina-kutya]-to
       get.up.early-do-must-COMP
       iwa-re-ta   sooda].
       tell-PASS-PAST   seem
       Zeniemoto-no
       former-head.of.school-GEN
       sekkyokusa-o   mago-ga
       positiveness-ACC   grandson-NOM
       monogatattei-te,…
       give.evidence-and
       'I heard that the new head of school was told by his son to get up early and pull up water from the well so as not to be preceded by his father. The grandson gave evidence of the former head of school's positive attitude…'
       (BCCWJ)

In order to illustrate the process of calculation of RD, let us measure the RD of *zen-iemoto* 'the former head of school'. The first step is to check the antecedent of *zen-iemoto*. Here, it is *ojiichan* 'grandfather' because it refers to the same person that *zen-iemoto* does. The second step is to calculate the clause boundaries between the target referent *zen-iemoto* and its antecedent *ojiichan*. In this study, the linear order of arguments determines RD of a referent. Following this approach, the RD

of *zen-iemoto* is 2. Here, zero subject intervenes between *zen-iemoto* and *ojiichan*.

3.2.5.2. Adjacent Predicates
V$_1$-*te*-V$_2$ form is basically categorized into the same clause, but when V$_1$ and V$_2$ have different subjects, each verb is regarded as belonging to an independent clause (Shimojō 2005: 57-8).

(7) a. kyanberu-no suupukan kat-te-ki-te
       Cambell-LK soup.can buy-TE-come-and
       '(I) bought a Cambell soup can (and came).'
    b. dorai-no-yatsu-o   tomodachi-ga
       dry-LK-one-ACC   friends-NOM
       motte-te (S) (O)   karite
       have-TE (I) (it)   borrow-and
       'A friend had dry (basil) and (I) borrowed it' (it).'
       (Shimojō 2005: 57-8)

For example, in (7a), the linked verb *kat-te-kite* 'buy-TE-come-and' share the zero subject 'I'. Thus, the V$_1$-*te*-V$_2$ form belongs to the same clause. In contrast, in (7b), V$_1$ and V$_2$ have different subjects. In other words, V$_1$ *motte* 'have' forms a nexus with *tomodachi* 'friend' and V$_2$ *karite* 'borrow' forms a nexus with the zero subject 'I'. In this case, both V$_1$ and V$_2$ are considered to constitute an independent clause because they do not share the same subject.

3.2.5.3. Back-channel feedback
Generally speaking, back-channel feedback such as *soo* 'indeed' and *un* 'yeah' are propositionally empty and are given by the hearer while speaker is holding the conversational turn (Shimojō 2005: 58). They are considered to be dependent on another clause and do not form an independent clause.

3.2.5.4. Copula
Copula expressions such as *da* and *dearu* are regarded as predicates and hence they head independent clauses.

3.2.5.5. Proposition
The method for determining RD has been developed for calculating the discourse status of a

referent (Givōn 1983, 1994). Proposition is not included in this method because it is not a referent itself but a relationship between referents. Instead of directly calculating the RD of a proposition, I count the RDs of the related referents. In my approach, the RD of the proposition is the least value of the referents pertinent to that proposition. For instance, in (8b), the scrambled object is the proposition *Hänsel-ga naka-ni hai-routosuru* 'that Hänsel is trying to come in it', which includes the referents *Hänsel* and *candy house*. Therefore, this proposition has the two related referents *Hänsel* and *candy house*. In this study, the RDs of both *Hänsel* and *candy house* are calculated. Note that the head of the scrambled object is nominalizer *no* but it is anchored by *Hänsel* and *candy house*. Thus, the RD of the scrambled object is replaced by the anchoring expression's RD and its RD is 1.

(8) a. okashinoie-ga   aru-node
      candy.house-NOM   be-because
      hutari-wa   hidoku   bikkurisuru
      two.person-TOP   very   surprised
      'Since there is a candy house, the two are
       very surprised.'

     b. Hänsel-ga   naka-ni
      Hänsel-NOM   inside-LOC
      hai-routosuru-no-o   Gretel-ga
      come-try.to.do-NMZ-ACC   Gretel-NOM
      togameru
      blame.for
      'Gretel berates Hänsel for trying to enter.'
                          (BCCWJ)

3.2.5.6. Movement Verbs
Movement verbs may affect the word order choice because Saeki (1960) points out that location tends to precede subject independently of information structure. Hence, locative objects placed in the sentence initial position are eliminated from my analysis.

3.3. Results
I analyzed 3273 examples from BCCWJ. Table 1 summarizes the distributions of scrambled objects from the viewpoint of RD. This table has demonstrated that 2676 examples have an

antecedent while 597 examples do not. Hence, 81.76% of objects in OSV are discourse-old information.

Table 1: Tokens of scrambled objects in terms of RD

RD	Number (%)
1	1724 (52.67%)
2	368 (11.24%)
3	194 (5.93%)
4	102 (3.12%)
5	61 (1.86%)
6	49 (1.50%)
7	34 (1.04%)
8	37 (1.13%)
9	19 (0.58%)
10	12 (0.37%)
11	14 (0.43%)
12	15 (0.46%)
13	8 (0.24%)
14	5 (0.15%)
15	5 (0.15%)
16	4 (0.12%)
17	5 (0.15%)
18	10 (0.31%)
19	10 (0.31%)
20+[2]	597 (18.24%)
Total	3273 (100%)

Table 2 is the summary of the distributions of scrambled objects in terms of *bunsetsu*. Recall that the subject in OSV is always 1 *bunsetsu* due to my design. Thus, more than one *bunsetsu* in Table 2 means the scrambled object is longer than its subject from the viewpoint of *bunsetsu*. Hence, heaviness correlates with scrambled objects in about 75.95% of examples, where the object is longer than one *bunsetsu*. However, there are many short scrambled objects in two *bunsetsu* due to the characteristics of *bunsetsu*. Although a demonstrative plus a NP constitutes two *bunsetsu*, it can be very short if the NP is short e.g. *sono-imi* 'that meaning' and *sono-hon* 'that book'. Thus, I

---

[2] 20+ includes the examples that have no antecedent.

counted the number of demonstratives plus NP that are short. Here, a 'short' NP means less than three characters. As a result, 149 of the two *bunsetsu* examples are short. Hence, they should be excluded from the heavy examples. Therefore, it is more appropriate to conclude that heaviness accounts for 71.40 % of the examples, which is the total ratio of 'real' heavy objects.

Table 2: The length of the objects in terms of *Bunsetsu*

Bunsetsu	Number (%)
1	787 (24.05%)
2	1028 (31.41%)
3	564 (17.23%)
4	379 (11.58%)
5	230 (7.03%)
6	109 (3.33%)
7	61 (1.86%)
8	44 (1.34%)
9	16 (0.49%)
10+[3]	55 (1.68%)
Total	3273

Next, Pearson correlation test was conducted between RD and *bunsetsu* in order to see if there is a correlation between givenness and heaviness. This analysis is based on the raw RD and *bunsetsu*. Consequently, it was revealed that there was no correlation between givenness and heaviness ($r = -.05$, $p<.01$). Thus, RD of the scrambled object is independent of its length.

3.4. Discussion
Generally speaking, the corpus analysis has demonstrated that OSV in Japanese is sensitive to discourse-old information. However, there are many counterexamples for the explanation that OSV is chosen when the scrambled object is discourse-old information. The first question I should ask is whether they are real counterexamples or not. In the following, I will point out that some counterexamples arise due to

---

[3] 10+ includes 10 and more than 10 *bunsetsu*.

weak points in my methods. Firstly, a sequence of same-reference NPs is called an appositive phrase, and such phrases are regularly discourse-old. In my approach, the direct object in (9) is regarded as completely new information because it has no antecedent in the preceding context. However, this example can be explained by supposing that the head of the scrambled object *enmoku* 'program' is activated by *nanatsumen* 'Seven Masks'. Thus, although the RD of *enmoku* 'program' is 20, it is not completely new information. Rather, it is possible that the NP *nanatsumen* 'Seven Masks' is introduced to the discourse in order to make the scrambled object given information. Thus, this type of example is not a crucial counterexample to my hypothesis.

(9) nanatsumen   nijuusuunen
     Seven.Masks   over.20.years
     enji-rarete-inai   enmoku-o
     perform-PASS-NEG   program-ACC
     Ebizoo-ga   aratana-kousoo-de
     Ebizoo-NOM   new-conception-with
     hukkatsu-sase-ta-toiu.
     revive-CAUS-PAST-seem
     'I heard that Ebizoo revived with a new conception the program called Seven Masks, which had not been performed for over 20 years.'

                                    (BCCWJ)

Secondly, let us look at scrambled 1st and 2nd persons. It has been said that interlocutors are conscious of each other (Chafe, 1987: 26; 1994:79). Thus, it is not too much to say that 1st and 2nd persons are permanently given information. In (10), the scrambled object *bokutachi-no-idokoro* 'our whereabouts' includes 1st person plural *bokutachi* 'we'. Although *bokutachi* 'we' has not been referred to in the previous discourse, it is given information because it is 1st person plural form. Hence, the scrambled direct object *bokutachi-no-idokoro* 'our whereabouts' as a whole can be regarded as given information.

(10) syainsyou-ga   haitteirun-dakara,

company.ID.card-NOM   have-because
sore-ga   tegakari-ni-nari,
it-NOM   clue-DAT-become
bokutachi-no-idokoro-o   keisatsu-ga
we-GEN-whereabouts-ACC   police-NOM
mitsukete-kureru-kamosirenai-yo
find-EMP-may-FP
'Since (my wallet) has my company ID
card, it can be the clue to our whereabouts
and police may find us.'

(BCCWJ)

Thirdly, some scrambled objects are semi-activated. Chafe (1987: 25) states that 'a semi-active concept is one that is in a person's peripheral consciousness, a concept of which a person has a background awareness, but which is not being directly focused on'. Furthermore, Chafe (1994: 86) states that a semi-active referent 'may be in the semi-active rather than new referents. It may be a referent that (a) was active at an earlier time in the discourse, (b) is directly associated with an idea that is or was active in the discourse, or (c) is associated with the nonlinguistic environment of the conversation and has for that reason been peripherally active but not directly focused on'. Note that RD can process type (a) and some parts of (b), but cannot deal with type (c). This is because RD counts on explicitly mentioned linguistic expressions. In (11), the scrambled object has no direct antecedent, but is semi-activated by non-linguistic context. *Kono-hon* 'this book' is a linguistically new referent because it has no antecedent, but it refers to the book a reader is reading now. The physical existence of 'this book' is a non-linguistic context.

(11)   kono-hon-o   dokusya-ga
       this-book-ACC   reader-NOM
       tenisuru-koroniwa...
       get-by.the.time
       'By the time readers get this book…'

(BCCWJ)

Next, I have demonstrated that heaviness correlates with 71.4% of the scrambled objects in OSV order examples. This supports Yamashita (2002) who observed that heaviness gave an explanation of about 74% (fourteen out of nineteen) of scrambled sentences. However, it is conceivable that heaviness depends on givenness, and vice versa. If such a tendency is universal in OSV as a whole, it is economical to use only one concept in order to explain the usage of scrambling. Therefore, I checked the correlation coefficient between givenness and heaviness. Consequently, it has been revealed that there is no correlation between them. Therefore, we can conclude that givenness and heaviness independently influence word order choices in Japanese. It is necessary to take both concepts into account in order to explain the function of OSV order.

Another point is that givenness correlates with 81.76% of scrambled sentences in my data. This result is opposed to Yamashita (2002) stating that scrambling is unrelated to information structure. In her data set, only 36.8% (seven out of nineteen) of objects are given information. Recall that her data set includes all kinds of scrambling: long-distance scrambling, short-distance scrambling, and VP-internal scrambling. On the other hand, my data set contains only OSV, which constitutes short-distance scrambling. Thus, it is conceivable that the strength of givenness effects depends on scrambling types. Let us expand upon this logic. With regard to OSV order, I have shown using a corpus that givenness correlates with the scrambled object. In contrast, with regard to S-DO-IO-V order, Ferreira and Yoshita (2003) observed that there was no interaction between given-new ordering and scrambling in sentence production. In other words, givenness has no strong influence on the choice of S-DO-IO-V word order. Why is there such a difference between OSV and S-DO-IO-V? One explanation relies on the mapping between information structure and syntactic structure. Note that OSV is VP-external scrambling whereas S-DO-IO-V is VP-internal scrambling. According to Rizzi (1997), sentence-initial position is related to discourse function. He supposes that left periphery architecture (CP-zone) is used to express the interfaces between syntactic structure and information structure. As shown in (12), the left periphery consists of many kinds of discourse-

related projections. Considering scrambling from the viewpoint of the left periphery, scrambled direct objects in OSV are considered to be related to information structure.

(12) Basic Structure of CP-zone and TP-zone
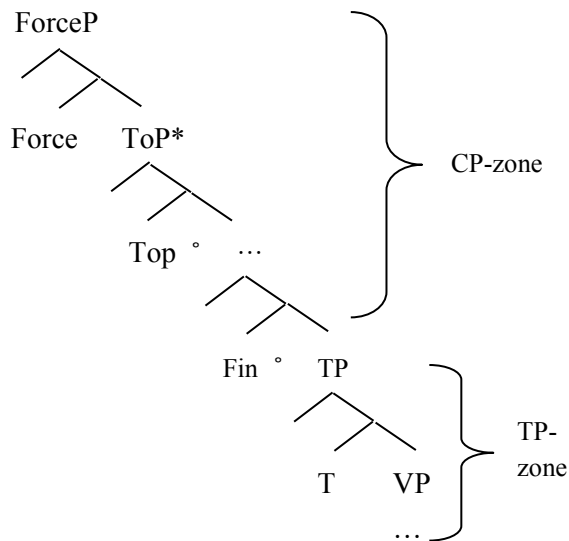
This is because sentence-initial position can have a relationship with information-related projections in the CP-zone. In particular, it may have a strong relationship with TOPIC projection, which seems to be pertinent to given information. In contrast, direct objects in S-DO-IO-V seem to be unrelated to information structure because there are no projections for information structure within the TP-zone. The TP-zone expresses only propositional meaning and information structure is not reflected in any projections within the TP-zone. In sum, givenness seems to have strong effects on OSV but have weak or no effects on S-DO-IO-V. This difference may be explained by the existence of discourse projections in the sentence-initial position. If this hypothesis is on the right track, long-distance scrambling will also be strongly influenced by information structure due to its position.

Numerous studies have shown that heaviness is an important factor for word order changes (Hawkins, 1994; Yamashita, 2002; Yamashita & Chang, 2001). The motivation for this phenomenon is to reduce processing cost, being unrelated to information structure. Hence, heaviness seems to have an influence on both VP-external and VP-internal scrambling. However, the strength of this effect may vary according to the informational status of scrambled constituents. In processing, when there are two competing factors, the strength of one factor becomes strong when the other factor is weak (Arnold et al., 2000; Trueswell and Tanenhaus, 1994). Applying this rule to heaviness and givenness, the effect of heaviness gets strong when that of givenness is weak, and vice versa. In fact, Yamashita (2002) observed that heaviness had a strong influence on VP-internal scrambling, although her data set was very small. Taken together, heaviness seems to have stronger effects on VP-internal scrambling than on VP-external scrambling. Conversely, givenness has greater effects on VP-external scrambling than on VP-internal scrambling. There might be such a complementary distribution between givenness and heaviness.

4. Conclusion

In this study, I have demonstrated that givenness has an influence on OSV order, being independent of heaviness. Specifically, it has been shown that the discourse-status of a scrambled object is important for the usage of OSV; OSV is preferable when O is discourse-old information. However, these conclusions contradict Yamashita (2002) claiming that information structure is not crucial for scrambling. This difference must derive from the data difference; her analysis includes VP-internal and VP-external scrambling while the scope of my analysis is only VP-external scrambling. Therefore, I propose that givenness has a strong effect on VP-external scrambling but a weak effect on VP-internal scrambling. The motivation for this analysis is based on Rizzi (1997)'s left periphery: sentence-initial position is related to information structure due to discourse-related projections. On the other hand, VP-internal scrambling seems to be unrelated to information structure because there are no discourse-related projections within the VP-internal zone. Moreover, since heaviness has no relationship with

information structure, it seems to have effects on both VP-external and VP-internal scrambling. However, heaviness may have greater effects on VP-internal scrambling than on VP-external scrambling. In order to examine the validity of these hypotheses, further corpus data are needed. Specifically, it is necessary to check the heaviness effects and the givenness effects on S-DO-IO-V.

Acknowledgement

I am deeply grateful to Dr. Stephen Wright Horn, Einar Andreas Helgason, Dr Tohru Seraku, and Prof. Hitosi Gotoo for their help.

References

Chafe, W. (1976). Givenness, Contrastiveness, Definiteness, Subjects, Topics and Points of View. In Charles N. (ed.), *Subjects and Topic*, 25-55. New York: Academic Press.

Chafe, W. (1987). Cognitive constraints and information flow. In Tomlin, R. (ed.), *Coherence and grounding in discourse*, 21-55. Amsterdam: John Benjamins Publishing.

Chafe, W. (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press.

Cooreman, A. (1992). The pragmatics of word order variation in Chamorro narrative text. In Doris, P. (ed.), *Pragmatics of word order flexibility*, 243-64. Oregon: John Benjamins Publishing.

Ferreira, V. S. & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of psycholinguistic research*, *32*: 669-692.

Givón, T. (1983) Topic continuity in discourse: An introduction. In Givón T. (ed.), *Topic Continuity in Discourse*, 5-41. Amsterdam: John Benjamins Publishing.

Givón, T. (1994) The pragamatics of de-transitive voice: Functional and typological aspects of inversion. In Givón T. (ed.) *Voice and inversion*, 3-47. Amsterdam: John Benjamins Publishing.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Ishii, Y. (2001). Presuppositional effects of scrambling reconsidered. In Inoue, K. and Hasegawa, N. (eds), *Linguistics and Interdisciplinary Research: Proceedings of the COE International Symposium*. Center of Excellence in Linguistics, 79-101. Tokyo: Kanda University of International Studies.

Kuno,S. (1978) *Danwa-no Bunpō* [Grammar of Discourse]. Tokyo: Taishūkan

Lambrecht, K. (1996) *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Publishing.

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koizo, H., Yamaguchi, M., Tanaka, M. & Den, Y. (2008). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, 48: 345-71.

Miyagawa, S. (2010). *Why Agree? Why Move? Unifying Agreement-based and Discourse-configurational Languages*, Cambridge: The MIT Press.

Rizzi, L. (1997). The fine structure of the left periphery. In Haegeman, L. (ed.), *Elements of grammar*, 281-337. Dordrecht: Kluwer Academic Publisher.

Saeki, T. (1960) Gendaibun ni okeru gojun no keikou. Iwayuru hogo no baai [Tendency of Word Order in Modern Japanese- Focusing on Complements]. *Gengoseikatsu*, 111: 56-63.

Saito, M. (1985) *Some Asymmetries in Japanese and their Theoretical Implications*. Doctoral dissertation, MIT.

Saito, M. (2009) Optional A-scrambling. *Japanese/Korean Linguistics*, 16: 4-63.

Shimojō, M. (2005). *Argument encoding in Japanese conversation*. Hampshire and New York: Palgrave Macmillan.

Vallduví, E. & Elisabet Engdahl. (1996) The linguistic realization of information packaging. *Linguistics*, 34: 459-520.

Yamashita, H., & Chang, F. (2001). Long before short preference in the production of a head-final language. *Cognition*, 81: B45–B55.

Yamashita, H. (2002). Scrambled sentences in Japanese: Linguistic properties and motivations for production. *TEXT-THE HAGUE THEN AMSTERDAM THEN BERLIN*, *22*: 597-634.