Unscrambling Some Misconceptions:  
A Comment on Koizumi and Tamaoka (2004)

Edson T. Miyamoto (University of Tsukuba)  
Michiko Nakamura (University of Hawai‘i)

Key words:  word order, scrambling, double object constructions,  
processing, frequency

Koizumi and Tamaoka (2004a) (KT hereafter) reported an experiment addressing the controversy surrounding the word order of double objects constructions in Japanese (Hoji, 1985; Matsuoka, 2003; Miyagawa, 1997). Although their conclusions may turn out to be correct, we argue that their conceptual justifications are problematic and more experimental work is needed, especially in relation to Matsuoka’s proposal.

We are particularly concerned with KT’s claim that scrambled word orders take longer to process because their syntactic representations are more complex compared to canonical orders. Such metrics based on competence models fail to address critical issues in performance raised in the 1970’s and later (Fodor et al., 1974). One problem is that in sentence processing we must consider incomplete sentences, which are often ambiguous, whereas syntactic theories have the luxury of dealing only with complete structures. Therefore, the mapping from competence theories to performance is not as straightforward as KT imply. In this commentary, we discuss KT’s claims within an incremental model of sentence processing and consider the kinds of hurdles one must overcome before empirical results can distinguish between competing syntactic proposals.

Despite our misgivings about some aspects of KT’s work, we believe
that their effort to bring syntax and language processing together is the kind of interdisciplinary work needed to attain a detailed understanding of human cognition, and we hope that our comments will clarify some issues regarding their contribution.

1. Background

Native speakers of Japanese are faster to judge whether a transitive sentence makes sense if the argument NPs are presented in canonical order, ga-o (nominative-accusative), compared to when they are scrambled, o-ga (Chujo, 1983). Recently, an impressive range of constructions has been investigated using this type of plausibility task (Koizumi and Tamaoka, 2004b; Muraoka et al., 2004; Tamaoka et al., 2005; Tamaoka et al., 2004). Among those, KT investigated the order of object NPs in ditransitive constructions. Regardless of the verb class (pass-type or show-type, Matsuoka, 2003), reaction times were faster when the two object NPs were shown in the ni-o (dative-accusative) order than in the o-ni order. KT concluded that the ni-o order is canonical for all ditransitive verbs, in accordance with Hoji (1985), but against Matsuoka (2003) and Miyagawa (1997).

The next section explores KT’s result in more detail by describing a model of how scrambled constituents are processed on-line (i.e., in real time, word by word). We then discuss some limitations of the plausibility task, and report frequency data on scrambling.

2. The incremental processing of filler-gap dependencies

For fronted wh-phrases (e.g., Stowe, 1986) and various other constructions, filler-gap dependencies have been claimed to be processed as follows. As soon as a filler (e.g., a wh-phrase, the nominal head of a postnominal relative clause) is detected, readers start looking for a gap and posit it as soon as grammatically permissible. Readers may prefer to insert the gap early in order to free cognitive resources such as working memory, which otherwise would be consumed in order to maintain the
gap-requirement active.

Syntactic complexity (as measured by number of merges, tree nodes, transformational rules) is not a factor. Nevertheless, the grammar plays a crucial role. First, it helps identify the filler and its need for a gap. Second, it circumscribes the possible positions where the gap can occur (Stowe, 1986, on subjacency and gap prediction).

2.1 Scrambling and filler-gap dependencies

In the following example (from Miyamoto and Takahashi, 2002), we illustrate how the filler-gap model can explain the processing of scrambled constituents in Japanese.

(1) Shokuin-ga ocha-o
employee-nom tea-acc

The NPs in (1) can be taken to be in their canonical positions as in a transitive construction. When kakarichoo-ni ‘manager-dat’ is read next, it becomes clear that this is likely to be the beginning of a ditransitive clause. Assuming for the moment that the ni-o order is always canonical, grammatical knowledge will indicate that the NP-acc was scrambled and requires a gap. The gap can be inserted immediately after the NP-dat, and because of working memory constraints, we assume that Japanese readers do so (note that the gap is inserted before the verb is seen; Aoshima, et al., 2004, for evidence).

(2) Shokuin-ga ocha-o kakarichoo-ni gap,
employee-nom tea-acc manager-dat

Readers slow down when they read the NP-dat in (2) (Miyamoto and Takahashi, 2002, Experiment 1; cf. Koso et al., 2004; Yamashita, 1997). However, this result alone does not guarantee that a gap is created at that point as other factors contribute to the slowdown in this case (e.g., the NP-dat is unexpected as readers prefer the NP-acc to be followed by a verb, Miyamoto and Takahashi, 2002). Work using more complex con-
structions confirms that a gap is likely to be inserted at that point (Miyamoto and Takahashi, 2004).

In this model, scrambled orders are harder to process because the filler-gap dependency requires extra cognitive resources to be processed, not because of syntactic complexity. Note that it is not necessarily true that all syntactically complex representations require more cognitive resources to be processed.

2.2 A word-by-word view of KT's data

We agree with KT that their results and earlier studies (Miyamoto and Takahashi, 2002, 2004) are difficult to explain under a syntactic model in which both ni-o and o-ni are canonical (Miyagawa, 1997; Miyagawa and Tsujioka, 2004). KT also claim that the longer latencies to the o-ni order for both classes of verbs in Matsuoka (2003) undermine his proposal that o-ni is canonical for pass-type verbs whereas ni-o is canonical for show-type verbs. But KT do not take into consideration potential ambiguities in mid-sentence in their items. (The same criticism applies to Miyamoto and Takahashi, 2004, the overall tendency they found for gap insertion after o-ni cannot be used against Matsuoka.) Example (3) (KT's example (7)) was intended as a pass-type item, but there is no guarantee that readers favor this interpretation after they see the first three NPs.

(3) John-ga hanataba-o Mary-ni
    John-nom flowers-acc Mary-dat

Within Matsuoka's model, whether a gap is created depends on the verb class. But crucially, the decision to insert the gap is made before the verb is seen; and the type of verb expected depends on various factors (e.g., whether the NPs are more plausible with a pass-type verb than a show-type verb). According to Matsuoka, a gap should be created if readers expect a show-type verb to follow.

(4) NP-ga NP,-o NP-ni gap, F-[expectation for a show-type verb]
However, when the pass-type verb *watasita* is seen next, the gap has to be removed (in Matsuoka’s classification, o-ni is the canonical order for pass-type verbs).

(5) \[ \text{NP-ga NP-o NP-ni gap, pass-type-verb.} \]

We are not aware of any work exploring the kind of reanalysis necessary to correct the mental representation for (5). But evaluation of Matsuoka’s model in KT’s experiment crucially depends on the cost of this reanalysis. Another factor that needs to be considered is how surprising (and consequently how costly) it is to process a pass-type verb when a show-type verb was initially expected. In order to provide a fair experimental assessment of Matsuoka’s model it would be necessary to create sentences for which the first three NPs can only be used with pass-type verbs or only with show-type verbs, so that temporary ambiguities are eliminated. Whether this can be done is unclear since the verbs in Matsuoka’s classes are not homogeneous in relation to their arguments, that is, the same sequence of NPs is usually compatible with either type of verb. This makes it rather difficult to provide experimental results for or against Matsuoka (2003).

The discussion above also casts doubts on results in acquisition that claim to favor Matsuoka’s proposal. Four-year-old children were more accurate to act out sentences in the o-ni order than in the ni-o order with pass-type verbs, but word order did not affect performance for show-type verbs (Isobe et al., 2004). The problem once more is that it is unclear which interpretation is favored before the verb is processed. Hence, predictions from syntactic models are again confounded by the ambiguity of the input sentence during comprehension. One can speculate that Matsuoka’s verb classes are affecting children’s performance not during comprehension, but later during recall in the act-out task.

Clearly, further studies are necessary, if for no other reason than that the acquisition study and KT are reaching opposite conclusions on Matsuoka’s proposal.
3. Methodological issues

Some problems in the methods adopted by KT need to be addressed in future work. Most crucially, the plausibility task only measures the total time to read the whole sentence and therefore it is not very informative. In general, recent work in sentence processing has measured behavioral or neurological responses time-locked with the processing of individual words in sentences. Even with such detailed measures, it is not always straightforward to interpret the results because competing explanations are often available. Thus, it should not come as a surprise that KT's plausibility task is even harder to interpret given its coarse temporal resolution.

For scrambling in particular, slowdowns in reading times tend to be small and have either not been detected (e.g., Yamashita, 1997) or have only reached statistical significance when complex structures were used (e.g., Mazuka et al., 2002; Miyamoto and Takahashi, 2002, 2004). Those studies used variations of self-paced reading, in which participants see one region of the sentence at a time and press a button in order to reveal the next region and hide the previous one. The method may seem artificial but in English it is well correlated with more naturalistic methods such as eye-tracking (Just et al., 1982). Although one could argue that the methodology is less suitable for head-final languages, to our knowledge there are no reports of discrepancies between self-paced reading and eye-tracking—the main difference is that the former is less sensitive than the latter. This contrast has been attested for scrambling: For simple scrambled transitive sentences, the second NP took longer to read in eye-tracking, but the difference was not statistically reliable in self-paced reading (Mazuka et al., 2002). Clearly, it is of interest to determine why differences for simple sentences seldom detected in self-paced reading (but detected in eye-tracking) are nevertheless picked up so consistently by the plausibility task.

Another issue that needs consideration is the influence of factors other than syntactic complexity. For example, it is well-known that fre-
quency affects lexical access, and such effects have been claimed to generalize to the sentence level (MacDonald et al., 1994). KT dismiss frequency as a confound for their hypothesis and claim that syntactic complexity most likely determines frequency of use. But this ignores the influence of non-syntactic factors (e.g., NP length) on the production of scrambling (Yamashita and Chang, 2001).

There is some tantalizing evidence that the outcome of the plausibility task is not always consonant with frequency. For example, the ni-ga order was found to be faster than ga-ni for potentials but slower for passives (Tamaoka et al., 2005). In general, the ga-ni order is more frequent than the ni-ga order, therefore one could claim that frequency is unable to explain these findings. We believe that the conclusion is premature. The frequency of ga-ni in general may affect the reading times for the NPs before the verb is read, but the reading times at the verb may depend on the frequency of each order with each type of predicate (i.e., we may need to count instances of ni-ga followed by potentials separate from those instances followed by passives, or perhaps just a NP-ga or NP-ni immediately preceding each type of predicate; animacy may also be an issue). All those effects at each point in the sentence are conflated if only the reading times for whole sentences are examined.

We are not claiming that the difficulty in processing scrambled orders comes from frequency alone. Our point is that, in order to make clear the contributions, if any, of syntactic complexity during processing, we should eliminate alternative explanations (such as frequency) by conducting corpus counts and completion studies, or by adopting experimental designs that can avoid frequency effects (Miyamoto and Takahashi, 2004, for an example). There is already some evidence that scrambled orders are rare (Kuno, 1973, pp. 353–354; Miyamoto and Takahashi, 2002; Yamashita, 2002). We report two corpus counts and a completion study to further evaluate the frequency of scrambled constituents.
4. Corpus count 1: Frequency of ditransitive sentences

The occurrences of the ditransitive verbs used by KT were extracted from 38,383 sentences of the Kyoto University Corpus (Kurohashi and Nagao, 1997), which is labelled with parts of speech and dependency relations (e.g., between NPs and predicates). The 305 instances with a NP-ni and a NP-o in the same clause will be referred to as the coarse-grained sentences. After a manual check removed irrelevant cases (e.g., ni-marked temporals), the remaining 152 cases (the fine-grained sentences) were classified as abstract (NP-o was an abstract entity), collocation (NP-o and verb formed a collocation), metaphor (the verb was used in a metaphorical sense), or as in KT (similar to the use in KT’s items). Those classes are important as they may determine whether the NP-o can be placed far from the verb.

Results are reported for both coarse and fine-grained sentences as frequency may accrue at a coarse level (e.g., readers lump together all ni phrases regardless of their meanings) or at a more fine-grained level (i.e., only semantically similar ni phrases are tallied together).

4.1 Results

Throughout the corpus, there were 1373 instances of the verbs used by KT. The statistical tests reported are for the verbs used in KT’s plausible items, of which there were fewer occurrences of show-type (190) than pass-type verbs (930). (The remaining 253 were occurrences of pass-type verbs in KT’s implausible items, implausible for short.)

Restricting to the instances that contain a NP-ni and a NP-o in the same clause, Tables 1 and 2 report the results for the coarse-grained and the fine-grained sentences.
Table 1: Ditransitive counts for the coarse-grained sentences

<table>
<thead>
<tr>
<th>Usage</th>
<th>Pass-type</th>
<th>Show-type</th>
<th>Implausible</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ni-o</td>
<td>o-•ni</td>
<td></td>
</tr>
<tr>
<td></td>
<td>147</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ni-o</td>
<td>o-•ni</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ni-o</td>
<td>o-•ni</td>
<td></td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Ditransitive counts for the fine-grained sentences classified according to use

<table>
<thead>
<tr>
<th>Usage</th>
<th>Pass-type</th>
<th>Show-type</th>
<th>Implausible</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ni-o</td>
<td>o-•ni</td>
<td></td>
</tr>
<tr>
<td>as in KT</td>
<td>22</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>metaphor</td>
<td>30</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>abstract</td>
<td>9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>collocation</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Subtotals</td>
<td>62</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

For the verb uses that were most similar to the ones in KT’s items (row as in KT in Table 2), the frequencies of the two word orders do not differ by much for pass-type verbs (ni-o 22, o-•ni 23) or for show-type verbs (ni-o 10, o-•ni 8). Thus, frequency in this narrow sense cannot explain the advantage for ni-o in KT’s study. However, the numbers on this row are small, and it is difficult to generalize based solely on them.

Overall, the ni-o order is more frequent (in Table 1, $\chi^2=94$, $P<0.0001$; in the subtotals row of Table 2, $\chi^2=8.58$, $P<0.01$). Thus, it is possible that Matsuoka’s classification is correct, but independent factors lead ni-o to be more frequent for both types of verbs yielding faster reaction times for this order in KT’s study. In other words, there may be a dissociation between frequency and grammar (i.e., canonicity in Matsuoka’s classification), and KT’s study may be reflecting the former rather than the latter.

One trend seems to confirm that the plausibility task is susceptible to frequency. In KT’s data, regardless of word order, reaction times are faster for pass-type than for show-type sentences (KT do not report statis-
tical tests, but the difference is large enough that is likely to be reliable). Similarly, our results indicate that KT's pass-type verbs are more frequent than their show-type verbs. This is confirmed by other data (Amano and Kondo, 1999/2000): The mean familiarity for nine of KT’s pass-type verbs is 5.993, whereas for seven of their show-type verbs it is 5.728, and the mean frequency in the Asahi Shinbun from 1985 to 1998 is larger for pass-type (26,443) than show-type (7,169). If the plausibility task is affected by verb frequency, as seems to be the case, then it may also be affected by the frequency of the word orders.

5. Corpus count 2: Frequency of transitive sentences

We examined a larger sample of accusative NPs and how often they are scrambled before a subject. All occurrences of sentence-initial accusative NPs (4,621) were automatically extracted from 38,383 sentences of the Kyoto University Corpus (Kurohashi and Nagao, 1997). The restriction to sentence-initial accusative NPs should not affect generality and was adopted to facilitate automatic extraction and minimize manual checks. Instances were classified as to whether the NP-o was followed by a subject in the same clause or not. All NP-wa and NP-ga were manually checked, and all except five turned out to be subjects (one was a corpus labelling mistake; in the other four, the NP-wa was not a subject, e.g., a locative headed by de ‘at’.) Other particles can also mark a subject; hence, 23 NPs ending in sae, sika, demo, koso, mo and made were manually checked. Four were subjects (marked with mo ‘as well’).

5.1 Results

Table 3 summarizes the results. Instances were separated depending on whether the accusative noun was modified by an adjective, demonstrative or relative clause (row complex) or not (bare, including noun-noun compounds). The column with subject lists the number of accusative NPs followed by an overt subject. The column no subject reports the instances of subject omission; those are further classified based on whether the
object and verb were adjacent (subcolumn adjacent) or some material intervened (not adjacent).

Table 3: Results for the NP-o frequency counts

<table>
<thead>
<tr>
<th></th>
<th>With subject</th>
<th>No subject</th>
<th>Subtotals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>adjacent</td>
<td>not adjacent</td>
</tr>
<tr>
<td>Bare</td>
<td>13</td>
<td>1173</td>
<td>307</td>
</tr>
<tr>
<td>Complex</td>
<td>75</td>
<td>2441</td>
<td>616</td>
</tr>
<tr>
<td>Subtotals</td>
<td>88 (1.9%)</td>
<td>4537 (98.1%)</td>
<td>1625 (100%)</td>
</tr>
</tbody>
</table>

Of the 4,625 instances of fronted accusative NPs, 88 (less than 2%) preceded a subject, the remaining 98% were missing the subject ($\chi^2=4365.4$, $p<0.0001$). In the majority of the no subject instances, the accusative NP was next to the predicate; but even when we eliminate those cases, the proportion of with subject corresponds to only 8.7%. In either case, a fronted accusative NP is highly unlikely to be scrambled prior to an overt subject.

6. Completion study for object NPs

To corroborate the newspaper data, a completion study with 32 native speakers of Japanese was conducted. Thirty-two sets of four conditions were created by crossing an animate noun-phrase (e.g., fukuoka shusshin-no shain ‘the employee from Fukuoka’) with one of four markers (accusative, dative, nominative or topic). Results for the nominative and topic NPs are not reported here (Miyamoto et al., 2005, for details). Four lists were created following a Latin Square design so that each NP appeared once in each list. Each participant saw one of the lists printed in random order and was asked to complete each NP plus marker into a sentence.

6.1 Results

As can be seen in Table 4, the subject was often omitted in the comple-
tions (*no subject* subtotals: 84.3% for NP-o, 80.2% for NP-ni) and was seldom produced overtly (*with subject* rows) although no prior context identifying the omitted subject was provided.

Table 4: Results for NP-o and for NP-ni subclassified by clause type (in the *no subject* category) or by marker on the subject (*with subject* category). (*Others* included: imperatives, relative clauses, clefts, adverbial clauses followed by the main clause).

<table>
<thead>
<tr>
<th>How the object NP was completed</th>
<th>NP-o</th>
<th>NP-ni</th>
</tr>
</thead>
<tbody>
<tr>
<td>No subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dropped in main clause</td>
<td>186</td>
<td>184</td>
</tr>
<tr>
<td>Others</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>Subtotals</td>
<td>210 (84.3%)</td>
<td>190 (80.2%)</td>
</tr>
<tr>
<td>With subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ga-marked subject</td>
<td>26</td>
<td>29</td>
</tr>
<tr>
<td>Wa-marked subject</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>No case marker</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subtotals</td>
<td>39 (15.7%)</td>
<td>47 (19.8%)</td>
</tr>
<tr>
<td>Total number of completions</td>
<td>249 (100%)</td>
<td>237 (100%)</td>
</tr>
</tbody>
</table>

To better determine the relative position of the object NP in the *no subject* instances, these are classified in Table 5 according to the constituent that immediately follows the object. One clear difference is that NP-o tends to be adjacent to the predicate (64.7%) and is rarely followed by another object (2.4%), whereas a NP-ni is equally likely to be followed by the verb (36.7%) or another object (33.3%; $\chi^2=82.47, P<0.0001$).

### 7. Discussion

The results indicate that scrambled object NPs are infrequent. Both the completion results and the corpus counts reveal a clear tendency for avoiding object NPs scrambled before a subject. The completion data also suggest that few instances of NP-o (2%) are followed by a NP-ni. This has direct consequences for the investigation of the processing of scrambling as it requires us to parcel out frequency effects before we can make claims about the influence of other factors (such as syntactic complexity). We
Table 5: Breakdown of the No subject instances in Table 4 according to the constituent that followed the object (NP-o or NP-ni). (Others include types that had seven or fewer instances for either type of object.)

<table>
<thead>
<tr>
<th>Intervening between object and verb</th>
<th>NP-o</th>
<th>NP-ni</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nil (adjacent)</td>
<td>161 (64.7%)</td>
<td>87 (36.7%)</td>
</tr>
<tr>
<td>Locative</td>
<td>14 (5.6%)</td>
<td>2 (0.8%)</td>
</tr>
<tr>
<td>NP-ni</td>
<td>5 (2.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>NP-o</td>
<td>1 (0.4%)</td>
<td>79 (33.3%)</td>
</tr>
<tr>
<td>Others</td>
<td>29 (11.6%)</td>
<td>22 (9.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>210 (84.3%)</td>
<td>190 (80.1%)</td>
</tr>
</tbody>
</table>

only reported frequencies for sequences of NPs according to their case markers, but as noted earlier, in some cases it may be necessary to consider the animacy of the NPs and the type of co-occurring verb (e.g., NP order frequency with potentials and with passives).

These frequency effects do not require sophisticated procedures such as keeping track of the probability of derivational rules of a grammar. It is enough to consider the contingency probability of X after seeing Y. For example, how likely is it that a NP-ni will follow after a sentence-initial NP-o is seen? Very unlikely according to the completion data (around 2% of the time, Table 5); thus, readers should be surprised if a NP-ni does come up. Anticipatory processes (e.g., anticipating the word most likely to continue the fragment being read) may be disrupted as readers are less able to prepare for what is coming next, leading to longer reading times. In short, the frequency effects considered here are some of the simplest that need to be factored out before experimental results can measure the effects of grammatical complexity.

By that we do not mean to claim that frequency is the main (let alone the only) factor behind processing difficulty in scrambling. But given the present results, it should be clear that any researcher attempting to make claims about processing difficulty in scrambling should consider ways of addressing frequency issues.
8. Conclusion

In order to examine the effects of syntactic complexity on processing, it is necessary to rule out the influence of independent factors and provide a detailed mapping between theories of competence and language processing. This paper discussed the filler-gap model for processing scrambled constituents in order to evaluate results in Koizumi and Tamaoka (2004). We also reported data indicating that frequency is a potent confounding factor and should be dealt accordingly. We believe that KT's results are worthy of further investigations, and we hope the present discussion will help determine some of the issues that need to be addressed in the future.

References

Unscrambling Some Misconceptions


Mainichi Shinbun (1995) CD-ROM.


かき混ぜ文の即時処理と出現頻度
——KOIZUMI and TAMAOKA (2004) へのコメント——

エジソン ミヤモト

（筑波大学）

中村 美智子

（ハワイ大学）

小泉・玉岡（2004）は、二重目的語を伴うかき混ぜ文と他の文に対する反応時間を比較し、結果の違いが二つの文の統語構造の違いによるものであると主張した。本稿では、新聞コーパスから得た頻度や被験者に行った文完成作業の結果をもとに、かき混ぜ文が比較的まれな現象であることを示し、かき混ぜ文に対する反応時間を扱う際、頻度の影響が問題となることを指摘する。また、日本語かき混ぜ文の即時処理モデルを例示し、小泉・玉岡が考慮すべき様々な問題を提示する。

（受理日 2004年12月16日 最終原稿受理日 2005年8月29日）