ONLINE COMPREHENSION OF SOV AND OSV SENTENCES IN TURKISH WITH A SUPPORTING CONTEXT*

BARiş KAHRAMAN AND YUKi HIROSE
The University of Tokyo / JSPS and The University of Tokyo

1 Introduction

In this study, we investigated the online comprehension of canonical (SOV) and non-canonical (OSV) sentences in Turkish through a self-paced reading task with supporting context. We will report that OSV sentences are not more difficult to comprehend than SOV sentences when a supporting context was provided, and the comprehension difficulty of non-canonical sentences in Turkish, like in Finnish, may be due to discourse-based factors rather than the complexity/frequency-based factors (Keiser & Trueswell, 2004).

Previous studies have shown that canonical sentences are generally easier to comprehend than their non-canonical counterparts in various languages (e.g., Sekerina, 2003; Koizumi & Tamaoka, 2004). Some studies attribute the comprehension difficulty of non-canonical sentences to their structural complexity and/or structural infrequency, because canonical sentences tend to be simpler and have higher frequency (e.g. Frazier & Flores d’Arcais, 1989; Hujanen, 1997). On the other hand, some studies attribute the processing difficulty of non-canonical sentences to discourse-based factors (Keiser and Trueswell, 2004).

For example, Keiser and Trueswell argue that given information tends to occur earlier in sentences, while new information tends to occur later in the sentences. They show that when the subject and object are used in a supportive context, non-canonical sentences (OVS) are no more difficult to process than canonical ones (SVO) in Finnish. Based on this result, Keiser and Trueswell argue that the processing difficulty of non-canonical sentences found in previous

---

* This research was supported by Japan Society for the Promotion of Science (JSPS) to the first author (Sentence processing of Japanese and Turkish as a first and second language; Project No: 25/03004). We would like to thank reviewers and audience of WAFL10 at MIT and CUNY 2014 at the Ohio State University for their constructive comments and suggestions. We also would like to thank students who voluntarily participated in the experiment at Çanakkale Onsekiz Mart University. All remaining errors and shortcomings are, of course, our own.
studies was due to violations of discourse factors, rather than complexity/frequency-based factors. However, in establishing discourse relationships, Keiser and Trueswell recycle nouns from the preceding contexts in their test sentences. We therefore cannot be sure whether their results are due to repeated noun benefit or familiarity with given subject and object nouns in the preceding context. More importantly, we also do not know whether the impact of the discourse can be generalized to other languages that use different word orders. To address these issues, we used Turkish SOV and OSV sentences with pronouns (discourse-old referents) and discourse-new referents.

In Turkish, Kuribayashi (2009) reported that SOV sentences were easier to comprehend than non-canonical sentences. However, since Kuribayashi did not provide any discourse context in his whole-sentence reading experiment, we cannot be sure whether his results are due to complexity/frequency-based factors or discourse-based factors. Therefore, the current study has two aims. The first aim is to examine the impact of the discourse (i.e., familiarity with given subject and object NPs) on the comprehension of SOV and OSV sentences in Turkish. The second aim is to examine whether the findings of Keiser and Trueswell in Finnish can apply to other languages when a pronoun is used as discourse-old referent.

2 The present study

To examine the impact of the information status of the subject and object nouns on the comprehension of SOV and OSV sentences, we conducted a self-paced reading experiment with 35 native Turkish speakers. One set of our test sentences consisted of four conditions, as shown in (1) and (2). Example (1) shows the preceding context and (2) shows the test sentences.

(1) **Preceding context**
Tren istasyonu-ndaki bilet satıcısı-ının ismi Vedat-tı.
Train station-at ticket seller-GEN name Vedat-PAST
‘The name of the ticket seller at the train station was Vedat.’

(2) a. **Canonical given subject – new object**
O Mine-ı ald-at-iyor diye istasyon amiri söyle-di
He Mine-ACC cheat on-PROG that station master say-PAST
‘The station master said that he cheats on Mine.’

b. **Canonical new subject – given object**
Mine o-nu ald-at-iyor diye istasyon amiri söyle-di
Mine he-ACC cheat on-PROG that station master say-PAST
‘The station master said that Mine cheats on him.’

c. **Non-canonical new object – given subject**
Mine-ı o-ald-at-iyor diye istasyon amiri söyle-di
Mine-ACC he cheat on-PROG that station master say-PAST
‘The station master said that he cheats on Mine.’

d. **Non-canonical given object – new subject**
3 Results

Prior to statistical analyses, we removed the two subjects and two items with incorrect response rates to the comprehension questions greater than 30%. We then removed reading times longer than 5000 milliseconds (ms) or shorter than 250 ms. Finally, we calculated 4 standard deviations (SD) for each region and replaced the data points exceeding 4 SDs from the mean with 4-SD boundary values. Since the word lengths were different in the first two regions, we also calculated the residual reading times and conducted statistical analyses on the residual reading times. The mean residual reading times of four conditions are shown in Figure 1.

Figure 1. Residual reading times of test sentences
The results of analysis of variance (ANOVA) showed that at the first word, the main effect of the information status of the nouns was significant in the item analysis \( (F_1 (1,32) = 3.84, p=0.059; F_2 (1,21) = 5.33, p=0.031) \). However, neither the main effect of word order \( (F_1 (1,32) = .03, p=0.87; F_2 (1,21) = .03, p=0.85) \) nor the interaction between the word order and information status were not significant \( (F_1 (1,32) = .44, p=0.51; F_2 (1,21) = .35, p=0.56) \). This shows that pronouns (given nouns) were read faster than the discourse-new nouns at the first word. At the second word, the main effects of the information \( (F_1 (1,32) = 12.26, p=0.0001; F_2 (1,21) = 40.84, p=0.0001) \), and the word order \( (F_1 (1,32) = 12.13, p=0.0001; F_2 (1,21) = 7.13, p=0.014) \) were both significant, but the interaction between the word order and information status was not significant \( (F_1 (1,32) = 0.19, p=0.66; F_2 (1,21) = 0.21, p=0.65) \). This shows that the embedded verbs after the new-given noun order are read faster than the verbs after the given-new noun order, irrespective of word order. At the fourth, fifth and sixth words, there was no significant main effect or interaction. At the sentence-final matrix verb position (i.e. *söyledi*), the main effect of the word order was significant in the subject analysis \( (F_1 (1,32) = 5.51, p=0.025; F_2 (1,21) = 3.06, p=0.095) \), but the main effect of information status \( (F_1 (1,32) = .61, p=0.45; F_2 (1,21) = 1.81, p=0.19) \), and the interaction between word order and information status of the nouns were not significant \( (F_1 (1,32) = 0.68, p=0.41; F_2 (1,21) = .94, p=0.34) \). These results show that the non-canonical sentences were actually read faster than the canonical sentences at the matrix verb position.

### 4 Discussion

In this section, we will primarily discuss the reading times observed at the subject and object NPs, the embedded verb and matrix verb, respectively.

At the first two words, where the subject and object NPs were presented, given-new noun order was easier to process than the new-given noun order, as in Finnish. Specifically, the initial positions of the sentence’s pronouns were read faster than the proper nouns, irrespective of their grammatical role. At first glance, one may think that this result may simply be due to the use of pronoun itself, rather than the information status of the nouns. If we only compare the reading time pattern of the first words, this possibility is plausible. However, at the second word, this pattern changed, and discourse-new nouns were read faster than the pronouns. This finding suggests that the processing ease of pronouns at the first word is due to the information status of subject and object NPs. The overall results observed at the first and second words confirm that given-new information order is also easier to process than new-given information order in Turkish, regardless of the surface word order, as in Finnish (Keiser and Trueswell, 2004).

At the embedded verb position, the new-given noun order was surprisingly read faster than given-new noun order, irrespective of whether they were in the canonical word order or not. This shows that canonical sentences are not necessarily easier to process than non-canonical sentences.
at the embedded verb position. This result is consistent with the findings of Özte, Marinis and Zeyrek, (2013), but differs from the previous findings reported in Turkish (Kuribayashi, 2009) and Finnish (Kaiser and Trueswell, 2004). Kuribayashi (2009) reported that canonical sentences are easier to comprehend than non-canonical sentences in Turkish. In his study, Kuribayashi did not provide any context prior to test sentences. In this study, we provided a context and manipulated the familiarity of given subject and object NPs in the test sentences. As a result, new-given noun order was easier to comprehend than given-noun order. This suggests that the use of context influences the processing of canonical and non-canonical sentences in Turkish, and the ease of comprehending canonical sentences may be due to lack of context in Kuribayashi (2009). Moreover, our result cannot be explained by complexity/frequency-based accounts, since they predict that canonical sentences should be read faster than non-canonical sentences due to the higher frequency and greater simplicity of canonical sentences in Turkish (Demiral, 2007).

In Finnish, Kaiser and Trueswell (2004) reported that when context is provided, sentences with given-new noun order are processed more easily than sentences with new-given noun order, irrespective of canonical word order. In this study, unlike Finnish, sentences with new-given noun order were easier to process than sentences with given-new noun order at the embedded verb position. We assume this result reflects the fact that the use of pronouns might have influenced the results. While Kaiser and Trueswell used the same nouns in the context and test sentences, we used third person singular person pronouns for the sake of avoiding repeated noun effects and naturalness in Turkish. When we look into observed reading time pattern, it is clear that the words immediately following the pronouns were read faster than the pronouns. This suggests that the use of pronouns might have sped up the reading times of embedded verbs in the test sentences. Moreover, different reading time patterns in Turkish and Finnish suggest that the use of discourse context influences the processing of canonical and non-canonical sentences, but the impact of discourse function on the sentence processing may have some language-specific properties too.

At the matrix verb position, surprisingly, non-canonical sentences were read faster than their canonical counterparts. Regarding this finding, we assume that the following possibility might have influenced the results: in the canonical sentences, sentence-initial subject nouns (i.e. o, Mine) might have been initially interpreted as the subject of the embedded verb. However, when the participants encounter the complementizer (diye), they need to change their simplex sentence interpretation to complex sentence. While doing this operation, the participants might have assumed that the sentence-initial subject NPs are argument of the matrix verb, and inserted a clause boundary before the object NPs (i.e., o-nu, Mine-yi). When the participants encountered the matrix subject (i.e., istasyon amiri (station master)) before the matrix verb, they realized that the sentence-initial subject NPs are not the argument of the matrix verb, and they might have changed their interpretation again, and then inserted the clause boundary in front of the sentence-initial position. In the case of non-canonical sentences, there was no such clause boundary ambiguity because scrambled subject-NPs cannot be interpreted as the argument of the matrix verb at the complementizer position. In other words, the participants do not need to change their initial interpretation in the non-canonical sentences. Therefore, this kind of clause boundary ambiguity and reinsertion of clause boundary might have caused longer reading times in the canonical sentences compared to non-canonical sentences.

Interestingly, there was also remarkable difference within the canonical sentences, and the matrix verb was read faster in the canonical new subject - given object condition than the canonical given subject new object condition. Regarding this result, two possibilities can be
considered. The first possibility is that the pronouns could have been interpreted as the argument of the embedded clause, while the proper nouns could have been interpreted as the argument of matrix clause. In other words, the participants needed to change their initial interpretation of proper nouns, whereas they did not need to change their interpretation of pronouns. Since, the burden of changing the initial interpretation is heavier for proper nouns than that for pronouns; the matrix verb might have been processed more easily in the *new subject – given object condition* than *given subject – new object condition*. The second possibility is that a binding ambiguity might have influenced the results. In the *new subject – given object condition*, either the embedded subject or the matrix subject can be interpreted as the antecedent of the discourse-old object noun. On the other hand, in the *given subject – new object condition*, there is no such binding ambiguity between the pronoun (discourse-old subject) and the matrix subject. Therefore, there might have been an asymmetry at the matrix verb in the canonical sentences. However, at this stage, we cannot distinguish between these possibilities. In a future study, we will attempt to explore this difference in Turkish.

Taken together, these results suggest that canonical sentences are not always easier to comprehend than non-canonical sentences, and the use of discourse context influences the processing of canonical and non-canonical sentences (Kaiser & Trueswell, 2004; Koizumi, 2010). However, our results suggest that the impact of discourse on the comprehension of canonical and non-canonical sentences may differ from language to language. Moreover, our study also suggests that the complexity/frequency-based accounts alone cannot solely explain the comprehension ease/difficulty of canonical and non-canonical sentences.

### 5 Conclusions and Future Directions

The current study had two aims. The first aim was to examine the impact of the discourse context, namely the role of given subject and object NPs on the comprehension of SOV and OSV sentences in Turkish. The results showed that when a context was provided, the new-given noun order was read faster than given-new noun order at the embedded verb position, irrespective of surface word order. At the matrix verb position, non-canonical sentences were read faster than the canonical sentences. Taken together, these results suggest that in Turkish, the discourse has an impact on the processing of canonical and non-canonical sentences, and canonical sentences are not always easier to process than non-canonical sentences. Moreover, these results also suggest that the complexity/frequency-based accounts alone cannot explain the processing ease/difficulty of canonical and non-canonical sentences.

The second aim was to examine whether the findings of Keiser and Trueswell (2004) in Finnish can apply to other languages when a pronoun is used as discourse-old referent. The results showed that, as in Finnish, the given-new noun order was easier to process than the new-given noun order, even when pronouns were used. On the other hand, at the embedded verb positions, this pattern was reversed, and new-given noun order was read faster than given-new noun order. Overall, our study suggests that the existence of discourse influences the processing of canonical and non-canonical sentences, but its impact on the sentence processing may be language-specific and differ from language to language.

In the current study, we only used pronouns as a discourse-old referent. As we argued above, the use of pronouns might have led to the present results. To examine the impact of noun-types on the processing of canonical and non-canonical sentences, we will conduct follow-up
experiments with various types of nouns. In this study, we did not set up any baseline condition. In order to examine how the existence of discourse context influences the processing of canonical and non-canonical sentences, we will also conduct reading experiments with and without a discourse context. Moreover, to explore how canonical and non-canonical sentences are used in the corpora, we need to conduct a corpus analysis. We leave these issues for future studies.

References


