Online comprehension of SOV and OSV sentences in Turkish with a supporting context

**Introduction:** In this study, we investigated the online comprehension of canonical (SOV) and non-canonical (OSV) sentences in Turkish using a self-paced reading paradigm. 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 (Keiser & Trueswell, 2004), may be due to discourse-based factors rather than the complexity/frequency-based factors (c.f. Kuribayashi, 2009).

Previous studies have shown that canonical sentences are generally easier to comprehend than their non-canonical counterparts in various languages. Some studies attribute the comprehension difficulty of non-canonical sentences to their structural complexity and/or structural infrequency because canonical sentences are less complex and have higher frequency (e.g. Frazier & Flores d’Arcais, 1989; Hyönä & 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 argued that old (given) information tends to occur early, and new information tends to occur later in the sentences. They showed that when the subject and object were mentioned in a supportive context, non-canonical sentences (OVS) were not more difficult to process than canonical ones (SVO) in Finnish. Based on this result, Keiser and Trueswell concluded that the processing difficulty of non-canonical sentences in the previous studies was due to violations of discourse factors, rather than the complexity/frequency-based factors. However, in establishing discourse relationships, Keiser and Trueswell repeated nouns from the preceding contexts in their test sentences. Therefore, we are not sure whether their results are due to repeated noun benefit or givenness of the 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 having different word orders.

In Turkish, Kuribayashi (2009) conducted a whole-sentence reading experiment and showed that canonical SOV sentences were easier to comprehend than non-canonical OSV sentences. However, since Kuribayashi did not provide any discourse context in his experiment, we are not sure whether his results are due to complexity/frequency-based factors or discourse-based factors (lack of context). Therefore, the current study aimed to examine the impact of the discourse on the comprehension of SOV and OSV sentences in Turkish, and explore whether the findings of Keiser and Trueswell in Finnish can be generalized to other languages, as well.

**Experiment:** We conducted a self-paced reading experiment with 35 native speakers of Turkish to examine the impact of the information status of the subject and object nouns on the comprehension of SOV and OSV sentences by manipulating the information status of the subject and the object noun. In the present study, using Latin Square design, we presented 24 target sentences with 48 filler items in random order. The target sentences were always preceded by a context as shown in (1), and embedded in the beginning of a complement clause, as shown in (2a)-(2d). To avoid unnaturalness and repeated name effects, we used pronouns when referring to given referents in the target sentences. By doing so, we can also provide a cross-linguistic evidence for the results of Keiser & Trueswell (2004), regarding the repeated name effects. If the results of Kuribayashi (2009) are simply due to complexity/frequency-based factors, the canonical sentences should be read faster than the non-canonical sentences because canonical sentences have higher frequency and less complexity in Turkish (Demiral, 2007). Moreover, this difference should be observed at embedded verb position (aldatyor) because the argument structures of the SOV and OSV sentences are first determined at the embedded verb position. On the other hand, if the results reported in Kuribayashi (2009) are due to lack of discourse, the comprehension difficulty of the embedded verb will be affected by the information status of the nouns, irrespective of word order.

**Test sentences:** (1) Preceding context, (2) Target sentences

(1) Tren istasyonu-n’daki bilet satıcısı’nın ismi Vedat’ti.
    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-yi aldat-iyor diye istasyon amiri söyle-di
   He Mine-ACC cheat-PROG that stationmaster say-PAST
   ‘The stationmaster said that he (Vedat) cheats Mine.’

   b. **Canonical – New subject / given object:**

   Mine o-nu aldat-iyor diye istasyon amiri söyle-di
   ‘The stationmaster said that Mine cheats him (Vedat),’

   c. **Non-canonical – New object / given subject:**

   Mine-yi o aldat-iyor diye istasyon amiri söyle-di
   Mine-ACC he cheat-PROG that stationmaster say-PAST
   ‘The stationmaster said that he (Vedat) cheats Mine.’

   d. **Non-canonical - Given object / new subject:**

   O-nu Mine aldat-iyor diye istasyon amiri söyle-di
   He-ACC Mine cheat-PROG that stationmaster say-PAST
   ‘The stationmaster said that Mine cheats him (Vedat).’

**Results and discussion:** The results of an ANOVA showed that the main effect of information status of the nouns was significant at the embedded verb (aldat-iyor) position ($F_1(1,32) = 5.78$, $p=0.022$; $F_2(1,21) = 5.51$, $p=0.029$). On the other hand, the main effect of word order and the interaction between word order and information status of the nouns were not significant. This shows that the verbs after the new-given noun order are read faster than the verbs after the given-new noun order, irrespective of word order. This may be due to fact that pre-verbal position is focus position in Turkish, and given nouns are easier to be focused than the discourse-new nouns. This result indicates that the comprehension ease/difficulty of canonical and non-canonical sentences is influenced by the information status of the nouns, rather than the word order per se. This result is inconsistent with the findings of Kuribayashi (2009), and suggests that non-canonical sentences are not more difficult to comprehend than canonical sentences in Turkish (Özge, Marinis and Zeyrek, 2013), when a supporting context is provided (Keiser and Trueswell, 2004). In the first two words, only the main effect of the information status of the nouns was significant. In the first word, given nouns (pronouns) were read faster than the discourse-new nouns ($F_1(1,32) = 3.84, p=0.059$; $F_2(1,21) = 5.33, p=0.031$). In the second word, on the other hand, discourse-new nouns were read faster than the given nouns ($F_1(1,32) = 12.26, p=0.0001; F_2(1,21) = 40.84, p=0.0001$). These results confirm that old-new information order is also easier to process than new-old information order in Turkish, regardless of the surface word order, as in Finnish (Keiser and Trueswell, 2004).

**Conclusion:** Overall, this study suggests that non-canonical sentences are not always more difficult to comprehend than the canonical ones, and that the comprehension difficulty of non-canonical sentences in the previous studies may indeed be due to violations of discourse demands. More importantly, complexity/frequency based accounts cannot explain the observed processing patterns in Turkish, while discourse based factors can explain. Thus the current study extends Keiser and Trueswell’s results for Finnish.

**Selected references:**


