On-Line Processing of Russian Scrambling Constructions: Evidence from Eye Movements During Listening

Irina A. Sekerina
(University of Pennsylvania)

1. Introduction

The central issue in current work on sentence processing is to explain how readers and listeners recover the linguistic structure of a sentence, and how they coordinate different types of constraints to resolve numerous local ambiguities that arise during on-line comprehension. There are two competing classes of models in sentence processing: modular and interactive. According to the modular approach, processes which construct syntactic structure of a sentence operate independently from other processes in comprehension, which are responsible for semantic and pragmatic interpretation of a sentence in discourse. In contrast, the interactive models assume that syntactic interpretation takes places with respect to a reader or listener’s knowledge of the contents of the prior discourse, which forms the context in which the sentence is processed. Thus, in the modular structure-based Garden-Path model (Kimball 1973; Frazier and Fodor 1978; De Vicenzi 1991; Frazier and Clifton 1996, among others), context does not influence the initial preferences in resolving local ambiguities as the syntactic structure is being built by the parser but only later gets incorporated into semantic interpretation. In the interactionist Constraint-Based model (MacDonald, Pearlmutter, and Seidenberg 1994; Tanenhaus and Trueswell 1995, among others), the parser is capable of coordinating the linguistic properties of the message with information from the context to determine processing commitments, on which it bases its ambiguity resolution strategies.
Traditionally, it has been difficult to observe contextual effects in studying language comprehension due to several factors. First, referential properties of the language are difficult to observe. Second, it is usually the case that language processing is investigated under artificial conditions, i.e., reading of either isolated sentences or short paragraphs. Finally, even when contextual information is presented, its use in reading is mediated by working memory, and only some subjects make effective use of contextual constraints (Just and Carpenter 1980). Recently, a new on-line technique has been developed which records the subject’s eye movements using a head-mounted eye-tracking system (Tanenhaus et al. 1996), making it possible to visually monitor the subject’s interpretation of the context while spoken language is being processed. Subjects’ eye movements are monitored as they respond to spoken instructions to move around objects on a table or flat shapes on a vertical board. This technique provides a new means of examining the moment-by-moment processes of subjects’ spoken language comprehension, in the relatively natural situation of acting upon spoken instructions. Section 2 presents a short background on research on English in which the head-mounted eye-tracking technique was employed for the first time (Tanenhaus et al. 1996). These studies showed that by monitoring eye movements of adults during listening, much can be inferred about the processes underlying language interpretation. Section 3 reports the results of an experimental study of Russian in which this technique has been used to examine on-line processing of Scrambling constructions in this language while establishing reference in temporarily ambiguous contexts.

2. Establishing Reference in English (Tanenhaus et. al. 1996)

Tanenhaus and colleagues used a head-mounted eye-tracking system very similar to the one used in the Russian experiment described below in Section 3. Subjects’ eye movements were recorded using a light-weight adjustable ISCAN eye-tracking visor
which looks like a helmet and consists of a monocle and two miniature cameras (see Figure 1). One camera records the visual environment from the perspective of the subject’s eye (the scene image), and the other camera records a close-up image of the eye. A computer analyzes the eye image in real-time, superimposing horizontal and vertical eye positions on the scene image. The scene image and the superimposed eye position, along with all auditory stimuli, are recorded on digital video tape.

Using this technique, Tanenhaus and colleagues (1996) studied how referents of definite nouns with adjectival modifiers are established in temporarily ambiguous visual contexts. The goal of the experiment was to find out whether the time necessary to identify such referents is affected by the point of disambiguation as determined by the characteristics of the potential referents in different visual displays. Five subjects listened to four critical commands illustrated in (1):

(1)  a. Touch the plain red square.
    b. Touch a plain blue triangle.
    c. Touch the starred yellow square.
    d. Touch the starred pink rectangle.
Each command was given in six types of visual displays (see Figure 2). Each display contained four blocks mounted vertically on a plastic board. The blocks differed in marking, color, and shape.

**NON-HOMOGENEOUS DISPLAY**

<table>
<thead>
<tr>
<th>EARLY</th>
<th>MID</th>
<th>LATE</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Blocks" /></td>
<td><img src="image2" alt="Blocks" /></td>
<td><img src="image3" alt="Blocks" /></td>
</tr>
</tbody>
</table>

**HOMOGENEOUS DISPLAY**

<table>
<thead>
<tr>
<th>EARLY</th>
<th>MID</th>
<th>LATE</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4" alt="Blocks" /></td>
<td><img src="image5" alt="Blocks" /></td>
<td><img src="image6" alt="Blocks" /></td>
</tr>
</tbody>
</table>

*Figure 2.* Types of Visual Displays in Tanenhaus et al. 1996, p.27

In non-homogeneous displays (top panel in Figure 2), the target ‘the plain red square’ differs from all other objects in the display in either marking, color, or shape. In homogeneous displays (bottom panel in Figure 2), other objects in the display are identical to the plain red square in either marking, color, or shape. The labels “Early”, “Mid”, and “Late” refer to the point of disambiguation in the instructions: early — on the first adjective, ‘plain’; mid-way — on the second adjective, ‘red’; and late — on the noun ‘square’.

A hypothesis referred to by Tanenhaus et al. as ‘the Incrementalist Hypothesis’ (p.18) was tested in this experiment. The Incrementalist Hypothesis states that when the context establishes
the set of likely referents, reference should be established immediately as the speech is being processed. Specifically, in the Late Disambiguation display, interpretation need not to wait until the head noun in the NP is identified.

The launch times for 72 (out of 120) trials on which the subjects’ first fixation was to the target (‘the plain red square’) were analyzed in a 2x3 ANOVA\(^1\) factorially combining Homogeneity of Display and Point of Disambiguation. Both effects were significant, showing a main effect of Homogeneity of Display, F(1,4)=13.03, p<.03, and a main effect of Point-of-Disambiguation, F(1,8)=5.94, p<.03. Thus, subjects were quicker to fixate on the target object in the homogeneous displays than in the non-homogeneous ones. The point of disambiguation, as determined by the instruction in conjunction with the display, influenced when eye movements occurred: faster in the early point of disambiguation display, slower in the mid one. Crucially, although eye movements to the target in the late point of disambiguation display occurred even later, they still preceded the onset of the head noun.

In sum, Tanenhaus et al. (1996) found that adults’ eye movements were closely time-locked with speech: subjects were typically launching eye movements to the intended referent within 300 msec of the onset of the disambiguating word, often before the end of that word. The resolution of reference was shown to involve a continuous integration of the linguistic information together with information present in the context.

---

\(^1\) ANOVA, or analysis of variance, is a statistical procedure used to determine whether means from two or more samples are drawn from populations with the same mean. F1 means that ANOVA is based on the subjects’ data, F2 — on the items’ data. The symbol “p” stands for probability and is statistically significant when it is less than 0.05. For basics of statistical analysis, see Ferguson and Takane 1989.
3. Establishing Reference in Russian Scrambling Constructions: An Eye Movement Study

3.1 Russian Word Order

Russian is a language which exhibits a rich morphological system of case marking. Subjects usually appear in the Nominative case, direct objects in the Accusative, and indirect objects in the Dative case. Thus, grammatical relations are reflected by the case marking and arguments can be freely ordered. In a three constituent sentence consisting of the subject, the verb, and the direct object all six combinations of these constituents are possible: SVO, SOV, OSV, OVS, VSO, and VSO.

Within the classical generative grammar framework (Chomsky 1986), different word orders in Russian are argued to be derived via movement referred to as Scrambling (Ross 1967). The treatment of Russian Scrambling within the framework of generative grammar (King 1995; Bailyn 1995), adopted as a background syntactic analysis for the purposes of this article, is quite different from the functional analysis of word order in Russian. In traditional Soviet, Russian and Prague School literature on word order in Russian and other Slavic languages, word order variants are related to the context in which they are appropriate. The context determines the bipartite division of every sentence into given and new information, the division known as the Functional Sentence Perspective (see Adamec 1966; Yokoyama 1986 for details). Bailyn (1995) discusses functional approaches to word order in Russian (see Chapter 3) and shows that while it is necessary to identify proper discourse conditions, as functional accounts do, it is not enough. Word order variations cannot violate principles of grammar, including movement constraints. In this sense, functional and generative grammar approaches to word order complement each other and are justified as legitimate ways of studying the same phenomenon.

A movement analysis of Scrambling in Russian presupposes that there is an underlying structure and order of the Russian clause and that phrases are scrambled from their base-generated
positions into landing positions higher in the clause to derive various surface word orders. Two principal types of clause-internal Scrambling can be identified in Russian: phrasal XP-Scrambling, and Split Scrambling. **XP-Scrambling**, illustrated in (2a), has been argued to represent an operation which moves a maximal projection (XP) from its base position to a landing position higher in the clause and which obeys restrictions of a familiar nature, for example, island constraints (see Bailyn 1995, esp. Chapter 2).

(2a)  
\begin{align*}
\text{Sobaku} & \quad \text{kupili} & \quad \text{na@i sosedi} & \quad \text{de@evo}.
\end{align*}
\text{dog}_{\text{ACC}} \quad \text{bought} \quad \text{our neighbors}_{\text{NOM}} \quad \text{cheaply}

‘Our neighbors bought the dog cheaply.’

(2b)  
\begin{align*}
\text{umnuju} & \quad \text{kupili} & \quad \text{na@i sosedi}. & \quad \text{sobaku}.
\end{align*}
\text{loud}_{\text{ACC}} \quad \text{bought} \quad \text{our neighbors}_{\text{NOM}} \quad \text{dog}_{\text{ACC}}

‘Our neighbors bought the loud dog.’

**Split Scrambling**, illustrated in (2b), is defined as an operation which breaks up NPs and PPs and moves one or both of their subparts into different positions in the sentence, thus deriving discontinuous constituents in which modifiers of all kinds are separated from the N head by other constituents in clause.

Word order variation in Russian makes it an interesting test ground for predictions of the theory of sentence processing. Although Russian is an SVO (see King 1995 for an alternative view), right-branching language (like English), it exhibits rich inflectional morphology with overt Case markers (like German) and free word order often thought of as discourse-oriented (like Japanese). On the other hand, unlike English, it is a Scrambling language, and unlike German and Japanese, it is not verb- or head-final. In addition, it allows Split Scrambling, i.e., discontinuous NPs and PPs. Split Scrambling constructions were used as experimental materials in the Russian experiment reported below.

This experiment was designed to investigate three specific questions. The first of these was whether there is evidence in Russian (as has been shown for English) for incremental use of con-
textually-defined constraints to establish referents for nouns modified by prenominal adjectives. The second goal was whether such prenominal adjective-plus-noun phrases are evaluated against general context, both linguistic and non-linguistic. How is reference for such NPs established? Could it be that it is not just on an incremental word-by-word basis, but perhaps on a finer, word-internal morphological subpart basis? The third goal of the experiment was to investigate whether the contrastive intonation facilitates the establishment of a referent in Split Scrambling constructions in Russian, thus making them less complex to process.

3.2 Method

Subjects. Sixteen volunteer subjects participated in this experiment, eight in each of the two versions of the experiment. All were undergraduate or graduate students at the University of Pennsylvania and native speakers of Russian who also spoke English as their second language. Typically, subjects took an hour and 15 minutes to complete the experiment.

Materials and Design. There were 24 experimental instructions involving a referential expression (noun) modified by an adjective. All the adjectives used in the experiment were color adjectives. Three types of visual displays were used in conjunction with instructions. Each display contained four shapes, two of which were the same shapes (see Figure 3).
(3) gives an example of the instructions which were designed in pairs and used in the displays illustrated in Figure 3, where (3a) contains XP-Scrambling while (3b) involves Split Scrambling:

(2) a. Po)alujsta, *krasnuju pti ku* polo)ite
    please    red\_ACC-FEM   bird\_ACC-FEM    put
    v poziciju 3.
    in position 3

    b. Po)alujsta, *krasnuju polo)ite pti ku*
    please    red\_ACC-FEM    put    bird\_ACC-FEM
    v poziciju 3.
    in position 3
    ‘Please, put the red bird in position 3.’

Informally, XP-Scrambling in (3a) requires an entire NP, the adjective and its head noun *krasnuju pti ku* ‘the red\_ACC-FEM bird\_ACC-FEM’, to be scrambled as a unit; in contrast, Split Scrambling in (3b) splits the adjective and the head noun by placing at least one other constituent (e.g., the verb *polo)ite* ‘put’) between them in the sur-
face word order. Usually, the semantic content of the utterance is not altered by either type of Scrambling.\(^2\)

For the Unambiguous display (the left panel in Figure 3), the target object \textit{krasnuju pti ku} (‘the red\textsubscript{ACC-FEM} bird\textsubscript{ACC-FEM}’) is the only red object, and the point of disambiguation at which there was sufficient lexical information to identify a single object as the target referent was right at the onset of the adjective ‘red’. For the Mid Point-of-Disambiguation display (the middle panel in Figure 3), the point of disambiguation was the ending on the adjective -\textit{uju} since the display contained another red object, \textit{krasnyj kvadrat} (‘the red\textsubscript{ACC-MASC} square\textsubscript{ACC-MASC}’)\(^3\) but this object’s grammatical gender was masculine in contrast to feminine gender of ‘the red bird’. Finally, for the Late Point-of-Disambiguation display (the right panel in Figure 3), the head noun was the point of disambiguation, because the display now contained two feminine red objects, the target red bird and the distractor red cup.

Each target instruction as in (3a-b) was followed by an additional distractor instruction referring to other objects in the same display. In addition, 16 fillers were interspersed with 24 experimental trials, resulting in a total of 40 trials. Experimental and filler trials were assembled to form two versions of the experiment in a fully counterbalanced design. The instructions and the type of display (Figure 3) were reflected in an experimental design factorially combining Scrambling Type (XP-Scrambling vs. Split

\(^2\) Although word order is free with respect to grammatical relations, it does alter the organization of a sentence on a communicative level, that is, its Theme/Rheme partition. The Theme is the starting point of the utterance and is often known to the listener or can be determined from the surrounding context. The Rheme tells the listener something about the Theme, carries the main communicative load of the utterance, and contains new information. In neutral speech the Theme precedes the Rheme (see Yokoyama 1986).

\(^3\) The Case marking on the phrase \textit{krasnyj kvadrat} (‘the red square’) is, in fact, ambiguous between masculine inanimate Accusative and Nominative cases, but this morphological ambiguity is not relevant for the present discussion.
Scrambling) and Point-of-Disambiguation (Unambiguous vs. Mid vs. Late), resulting in six conditions, with four target instructions per condition (2 x 3 x 4).

Procedure. Subjects were run individually, seated in front of a vertical board placed on a table. They were given spoken instructions to move various flat objects around on the board. Subjects were instructed to keep their eyes closed until they heard the word pojalústa (‘please’), which began the first (target) instruction in every trial. This was done so that subjects could not get acquainted with the display prior to the instructions. Their goal was simply to perform the instructions as naturally as possible. It is important to note here that there was no centrally located fixation cross, and subjects were free to look anywhere as soon as they opened their eyes.

Prior to the experiment a calibration procedure was performed for each subject. At the beginning of each trial, the four objects were verbally identified (without naming the colors) so that there was no confusion. The first (target) instruction in each trial started with the word pojalústa (‘please’), while the second one began with the word teper’ (‘now’). For instance, a trial would consist of the following instructions for the left panel in Figure 3:

(4) Pojalústa, krasnuju poloitite ptiku
    please redACC-FEM put birdACC-FEM
    v poziciju 3.
    in position 3
    Teper' poloitite jeltyj krug v poziciju 8.
    now put yellowACC circleACC in position 8
    ‘Please, put the red bird in position 3.
    Now put the yellow circle in position 8.’

All the instructions were produced by the experimenter live as they were read from the script during the experiment. Every effort was made to produce them with natural and consistent intonation.
While subjects followed the instructions to move objects around on the board, their eye movements were recorded using a light-weight ISCAN eye-tracking visor (Figure 1 above). The timing of eye movements relative to information in the speech stream was computed as follows: eye movement data for trials in which the initial fixation was to the correct object were analyzed from the video tape by identifying the beginnings of critical words for each trial, and noting the time lapse between the critical speech points and the onset of an eye movement to the intended object. Eye movement latencies were measured from the onset of the color adjective.

Digital video tapes of each subject’s scene were analyzed by hand, using slow motion and freeze frame viewing of the tapes. The auditory commands to move objects were recorded on the auditory channel of the tape and were also analyzed. A single scorer analyzed the tapes.

3.3 Results

Two types of data were analyzed in the experiment: percentages of looks at different shapes, and eye movement latencies. Since subjects were free to look anywhere on the display without fixating on a central point, the first look could be to either the target (the red bird in Figure 3 above), the color distractor (the red cup), or any other object including the shape distractor (the blue bird).

Table 1 shows the percentages of trials in which subjects looked at the color distractor at any point during the trial.

<table>
<thead>
<tr>
<th></th>
<th>MID DISAMBIG.</th>
<th>LATE DISAMBIG.</th>
</tr>
</thead>
<tbody>
<tr>
<td>XP-SCRAMBLING</td>
<td>68%</td>
<td>75%</td>
</tr>
<tr>
<td>SPLIT SCRAMBLING</td>
<td>66%</td>
<td>68%</td>
</tr>
</tbody>
</table>
The data indicate that subjects considered the color distractor (the competitor object), e.g., the red cup in Figure 3 above, in approximately 70% of all ambiguous trials, regardless of the point of disambiguation. Furthermore, it made no difference for subjects whether the instructions contained an XP-Scrambling (nonsplit) or Split Scrambling construction, since they tended to look at the distractor equally often in both. The analysis of variance supports these observations, since although there is a numerical difference in the Mid Point-of-Disambiguation Split condition, there are no statistically significant differences for either of the factors, i.e., Point-of-Disambiguation and Scrambling Type.

Table 2 shows percentages of trials in which subjects looked first at (a) Target, (b) Distractor, or (c) any Other object on the display:

<table>
<thead>
<tr>
<th></th>
<th>UNAMBIGUOUS</th>
<th>MID DISAMBIGUATION</th>
<th>LATE DISAMBIGUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>XP-SCRAMBLING:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td>86%</td>
<td>43%</td>
<td>36%</td>
</tr>
<tr>
<td>Distractor</td>
<td>—</td>
<td>52%</td>
<td>59%</td>
</tr>
<tr>
<td>Other</td>
<td>14%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td><strong>SPLIT SCRAMBLING:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td>100%</td>
<td>48%</td>
<td>55%</td>
</tr>
<tr>
<td>Distractor</td>
<td>—</td>
<td>45%</td>
<td>43%</td>
</tr>
<tr>
<td>Other</td>
<td>0%</td>
<td>7%</td>
<td>2%</td>
</tr>
</tbody>
</table>

These data show that subjects tended to look at either the Target or the Distractor equally often in the Mid and Late Point-of-
Disambiguation conditions. First looks to other objects on the display also occurred, but their percentages were extremely small.

For the percentages of first look to target, these data show a strong main effect of Scrambling Type, $F(2,24)=38.74$ $p<.001$, as well as a main effect of Point-of-Disambiguation, $F(1,12)=13.67$, $p<.005$, and no interaction of Point-of-Disambiguation and Scrambling Type, $F<1$. Figures 4 and 5 illustrate that subjects tended to launch their first eye movement to the Target in the Split Scrambling conditions (48% and 55%, respectively), but that in the XP-Scrambling Condition they looked first at the Distractor (52% and 59%).

![Figure 4](image_url)  
Figure 4. Percentages of First Looks at Target and Distractor in XP-Scrambling Condition
Figure 5. Percentages of First Looks at Target and Distractor in Split Scrambling Condition.

For eye movement latencies, summary data are presented in Figure 6 below:

Figure 6. Mean Eye Movement Latencies (msec) to Target
Only data from trials in which the subject made an eye movement to the Target were included in the analysis. Launch times were measured from the beginning of the adjective, e.g., from the beginning of the word *krasnuju* (‘red<sub>ACC-FEM</sub>’) in the examples in (3) above. In general, subjects initiated eye movements to the Target shortly after hearing the word in the instruction that disambiguated the target object from other objects in the display. They were the fastest in the Unambiguous condition, slower in the Mid Disambiguation, and the slowest in the Late Disambiguation conditions.

**Discussion.** The results show evidence for incremental processing with respect to a visually presented set of potential referents. In particular, the data indicate that nouns modified by adjectives are interpreted incrementally. Subjects considered the distractor object as soon as they started processing the adjective without waiting for the disambiguating information. This was evident even in the Split Scrambling conditions, where the head noun was separated from the modifying adjective by the verb. These results support the conclusion that the adjective-plus-noun phrase was interpreted incrementally on-line with respect to all the potential referents in the visual model.

The point of disambiguation, as determined by the instruction in conjunction with the display, clearly influenced when eye movements occurred. Eye movements to the target object began shortly after the disambiguating word. Thus, the position of the head noun which was manipulated in the experiment (adjacent to the adjective in the XP-Scrambling conditions and split by the verb in the Split Scrambling conditions) made no difference. Moreover, subjects were even faster in launching an eye movement to the target object in the Split conditions, usually fixating on the target prior to hearing the head noun in the instruction. This suggests that people have immediate access to the intonationally-marked contrast which was present in the Split conditions, as required by discourse requirements of contrastive function associated with Split Scrambling constructions.
4. Conclusions

As was the case with the English experiment (Tanenhaus et al. 1996), the results of the Russian experiment are consistent with the Incrementalist hypothesis: When the context presents the set of likely referents, reference is established immediately as the speech is being processed, without delay. They provide compelling evidence for a processing model in which linguistic expressions are undergoing continuous, moment-by-moment interpretation, with immediate mapping onto a discourse model and potential referents (MacDonald, Pearlmuter and Seidenberg 1994; Tanenhaus and Trueswell 1995). They highlight the incremental and referential nature of spoken language comprehension and demonstrate that linguistic and visual information are rapidly integrated in real-time processing as argued by the interactionist constraint-based model of sentence processing.

References


The Institute for Research in Cognitive Science
3401 Walnut Street, Suite 400A
Philadelphia, PA 19104
sekerina@linc.cis.upenn.edu