Russian sentence corpus:

Benchmark Measures of Eye movements in reading in Cyrillic

Anna Laurinavichyute (HSE, Moscow), Irina Sekerina (CSI CUNY, New York, Svetlana Alexeeva (SPbSU, St. Petersburg), Kristine Bagdasaryan (HSE, Moscow)

THE STUDY
Main goal: to create the first systematic corpus of eye movements in reading in Russian

Research Questions:
Q1. Do native speakers of Russian read texts as readers of other alphabetic languages (in particular, German) do concerning basic psycholinguistic variables (length, predictability, frequency)?
Q2. What about morphological characteristics (not tested previously)?

CORPUS DESIGN AND MATERIALS
Based on German PSC (Kliegl et al. 2004; 2006) protocol
Step 1. 144 target words (3*2*2 design): part of speech (adjective/noun/verb), length (3-4, 5-7, and 8-10 characters), and frequency (> 50 ipm or < 10 ipm)
Step 2. Sentence selection using RNC and acceptability norming (215 participants): 5-11 words in length, target word in the middle
Step 3. Collection of predictability norms (750 participants):
Predictability close word: Ваня раскрыл было рот, но понял, ?? (что)
Step 4. Eye-movement data collection (96 participants): Eyelink 1000 plus (with chinrest)

RESULTS
Q1. Corpus Reading. Descriptive statistics

<table>
<thead>
<tr>
<th>All words</th>
<th>Russian</th>
<th>German</th>
<th>All words</th>
<th>Russian Mean</th>
<th>German Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>N of sentences</td>
<td>144</td>
<td>144</td>
<td>FF</td>
<td>222 ms.</td>
<td>207 ms.</td>
</tr>
<tr>
<td>Sentence length Range: 5-13 words, M = 9</td>
<td>Range: 5-11 words, M = 7.9</td>
<td>SF</td>
<td>232 ms.</td>
<td>210 ms.</td>
<td></td>
</tr>
<tr>
<td>N of words</td>
<td>1362</td>
<td>1138</td>
<td>GD</td>
<td>231 ms.</td>
<td>241 ms.</td>
</tr>
<tr>
<td>Word length M = 5.6 let.</td>
<td>TT</td>
<td>283 ms.</td>
<td>245 ms.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictability M = 18%</td>
<td>PO (PSK)</td>
<td>17.8%</td>
<td>9.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guesses per word 20-151</td>
<td>83</td>
<td>P1</td>
<td>68.8%</td>
<td>74.1%</td>
<td></td>
</tr>
<tr>
<td>% of short words 35%</td>
<td>41%</td>
<td>P2*</td>
<td>4.9%</td>
<td>17.0%</td>
<td></td>
</tr>
<tr>
<td>% of LF words 61%</td>
<td>45%</td>
<td>RO</td>
<td>17.8%</td>
<td>12.5%</td>
<td></td>
</tr>
<tr>
<td>% of LF words 65%</td>
<td>66%</td>
<td>RG</td>
<td>7%</td>
<td>0.4%</td>
<td></td>
</tr>
</tbody>
</table>

Q2. Focus on morphology

C1. Part of Speech
Nouns are read faster than verbs in the GD and TT measures. Adjectives, adverbs, and function words did not differ from the verbs in any of the measures

C2. Morphosyntactic ambiguity
There was no difference in reading times between morphosyntactically ambiguous and unambiguous word forms in the Russian

C3. Base vs. Inflected word form
Inflected word forms took longer to read, but the effect was only significant in the TT measure

CONCLUSIONS
Q1. Gaze patterns are similar to the data reported for German (Kliegl et al. 2004; 2006) concerning basic word variables.
Q2. Part of speech and case influence eye-movement measures in reading, whereas morphosyntactic ambiguity does not.