Intercomprehension as a special mode of language use.

Certain degrees of mutual intelligibility between (closely) related languages in written and spoken modalities.

Existing and perceived similarities at different linguistics levels.


The parallel translations are split into seven distinct fragments, each fragment is aligned in a multiple alignment scheme, all corresponding syntactic units within a fragment match with one another.

The Russian-Bulgarian final fragment “And so the North Wind was obliged to confess that the Sun was the stronger of the two”: Ru: Танким образом северный ветер вынужден был увидеть буй.

BG: Таким обраzом северный ветер принуден да увидел са̀бун.

Ru: призывать что солнце светло e его

BG: призывать что слыньца e по-сильно от него

Corresponding units appearing in different sentence positions are highlighted in green, syntactic units without corresponding pairs are marked in red text.

Modeling intercomprehension among nine Slavic languages: Belarusian, Bulgarian, Croatian, Czech, Polish, Slovak, Slovene, Russian, and Ukrainian information-theoretically:

- RQ1: How syntactically distant are these nine Slavic languages from each other?
- RQ2: What asymmetries are predictable by means of adaptation surprisal between selected languages from phonetic and orthographic views?
- RQ3: What is the relation among the measures under study here?

### Measuring Methods

**SYNTACTIC DISTANCE** (Neeringa et al., 2017)

- The *InDel* distance (InDel) measures the average number of words (syntactic units) which are inserted or deleted in parallel sentences.
- The binary *movement distance* measures the average number of words that must be reordered in sentences of L1 in order to produce the word order of an equivalent sentence in L2.
- The linear *movement distance* measures the number of word positions a word from a sentence in L1 has moved compared to the corresponding word in an equivalent sentence in L2.

**ADAPTATION SURPRISAL** (Stenger, Avgustinova, and Martí, 2017)

- Adaptation surprisal, in particular Word Adaptation Surprisal (WAS), quantifies the degree of unexpectedness of a word form given a possibly related word form and set of transformation probabilities.

\[
WAS = \sum_{t=1}^{m} - \log_2 P(L_1|L_2)
\]

L1 refers to the \(i\)th character or sound in the native (decoder) language and L2 refers to the \(i\)th character or sound in the foreign (stimulus) language and

### Results: Syntactic Distances

<table>
<thead>
<tr>
<th>L1</th>
<th>L2</th>
<th>InDel</th>
<th>Linear Movement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Results: Adaptation Surprisal

<table>
<thead>
<tr>
<th>L1</th>
<th>L2</th>
<th>InDel</th>
<th>Linear Movement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Conclusions and Outlook

**RQ1:** The West Slavic languages are more similar to one another than the East and South Slavic languages.

**RQ2:** With only a few exceptions, the mean normalized adaptation surprisal is lower on the orthographic level than on the phonetic level. Czech and Slovak exhibit the least normalized adaptation surprisal with one another on average on both the orthographic and phonetic levels.

**RQ3:** High and significant correlations between the mean InDel distance and the mean orthographic (Pearson’s \(r = 0.86\); \(p < 0.001\)) and phonetic adaptation surprisal (Pearson’s \(r = 0.91\); \(p < 0.001\)), as well as between the two measures of adaptation surprisal (Pearson’s \(r = 0.91\); \(p < 0.001\)).

The exact prediction potential of measure methods will be validated with intelligibility scores obtained in web-based experiments among speakers of selected Slavic languages.

### References


Syntactic distances and surprisal between nine Slavic languages of the fable “The North Wind and the Sun”.

https://github.com/slavic-lab/LREC-2022-SynDist-Surprisal