Expression of information structure in Slavic: Experiments and modeling

Radek Šimík and Marta Wierzba
University of Potsdam / SFB 632

Kolloquium Slawistische Linguistik
Humboldt University Berlin, June 16th 2014
Research question

How is information structure expressed in West Slavic languages?

**Common view:** In Slavic languages information structure gets primarily expressed by word order. (In English, it is primarily expressed by prosody.)

**Our view:** The expression of information structure is primarily prosodic in Slavic languages (and only secondarily word-order-based).
Further issues

Optionality
- Syntax generates more outputs for a single meaning (example: focus in situ, focus fronting).
- Free (non-feature-driven) external and internal merge (Chomsky’s recent work).

Gradience
- Some outputs might be preferred over others (without those being unacceptable).
- We capture this by a post-syntactic Linear OT model.
- This model also allows for precise evaluation of (competing) hypotheses.
Basic notions: Givenness

**Givenness**

An expression $\alpha$ is **given** if and only if it is in the common ground that there is an expression $\beta$ in the preceding discourse such that

a. if $[\alpha] \in D_e$, then $[\alpha] = [\beta]$

b. if $[\alpha] \in D_{(\sigma,t)}$ (for any type $\sigma$), then $[\alpha] \supseteq [\beta]$

**Examples**

A: Did you speak to $[\beta \ Eddy]$?

B: Sorry, I don’t **know** $[\alpha \ Eddy]$.

A: Look, Jane is eating a $[\beta \ carrot]$.

B: Yeah, she **loves** $[\alpha \ vegetables]$.

Basic notions: Focus

**Focus**

An expression $\alpha$ in a background $[\beta \ldots \alpha \ldots]$ is **focused** if there is a salient set of alternatives to $\beta$ generated by replacing $\alpha$ with expressions of the same type.

**Examples**

A: Who surprised you the most yesterday?
B: $[\beta [\alpha \text{My first-year student}] \text{ surprised me}]$.

A: I’m afraid that the window remained open.
B: I think that the teacher $[\beta [\alpha \text{ closed}] \text{ the window}]$.

In English, givenness and focus are expressed prosodically.

**NUCLEAR STRESS RULE**
Place sentence stress on the rightmost constituent in a clause.

**STRESS FOCUS**
Place sentence stress on the focused constituent.

**STRESS GIVEN**
Do not place sentence stress on given constituents.

In Slavic, information structure is expressed by word order.

Example: given-left, focus-right

A: Kdo nakrmil psa?  
   ‘Who fed the dog?’

B: Psa nakrmil Milan.  
   dog.acc fed Milan.nom

cf. Milan fed the dog.
Two hypotheses of how this comes about

The syntactic approach
information structure $\leftrightarrow$ word order

The prosodic approach
information structure $\leftrightarrow$ prosody $\leftrightarrow$ word order
The syntactic hypothesis

A special constraint relating IS and word order is postulated.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Constraint Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathesius</td>
<td>1939</td>
<td>Theme &gt; Rheme (&quot;objective order&quot;)</td>
</tr>
<tr>
<td>Firbas</td>
<td>1971</td>
<td>less communicatively dynamic &gt; more communicatively dynamic</td>
</tr>
<tr>
<td>Kučerová</td>
<td>2007, 2012</td>
<td>given &gt; new</td>
</tr>
<tr>
<td>Biskup, Mykhaylyk</td>
<td>2011</td>
<td>[... given/old/specific ... [vP ... new/non-specific ...]]</td>
</tr>
</tbody>
</table>
Grammatical marking of givenness: In Czech (K 2007) and other Slavic languages (K 2012), a “G-operator” is inserted at LF. This operator marks everything in its scope (outside its c-command domain) as given. If a given expression is not marked by G as given, this results in a Maximize Presupposition violation. If a new expression is marked as given, this results in a presupposition failure.

Note: Kučerová’s theory is comparatively easy to test because the notion of givenness (and newness) can be very well defined and mapped onto data.
The prosodic hypothesis

Using the constraints assumed for English: STRESS FOCUS (SF), *STRESS GIVEN (*SG), NUCLEAR STRESS RULE (NSR)

Deriving the word order indirectly: via IS-prosody constraints

A: Kdo nakrmil psa?  
   'Who fed the dog?'  

B1: Psa nakrmil Milan  
B2: Milan nakrmil psa  
B3: Milan nakrmil psa  

Rationale: Slavic uses free word order to maximize the satisfaction of IS-prosody constraints.

The prosodic hypothesis

Using the constraints assumed for English: STRESS FOCUS (SF), *STRESS GIVEN (*SG), NUCLEAR STRESS RULE (NSR)

Deriving the word order indirectly: via IS-prosody constraints

A: Kdo nakrmil psa?  
   ’Who fed the dog?’

B₁: Psa nakrmil Milan.  
   ‘Milan fed the dog.’

The prosodic hypothesis

Using the constraints assumed for English: STRESS FOCUS (SF), *STRESS GIVEN (*SG), NUCLEAR STRESS RULE (NSR)

<table>
<thead>
<tr>
<th>Deriving the word order indirectly: via IS-prosody constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong>: Kdo nakrmil psa?</td>
</tr>
<tr>
<td>’Who fed the dog?’</td>
</tr>
<tr>
<td><strong>B₁</strong>: Psa nakrmil Milan.</td>
</tr>
<tr>
<td><strong>B₂</strong>: Milan nakrmil psa.</td>
</tr>
</tbody>
</table>

‘Milan fed the dog.’

The prosodic hypothesis

Using the constraints assumed for English: STRESS FOCUS (SF), *STRESS GIVEN (*SG), NUCLEAR STRESS RULE (NSR)

Deriving the word order indirectly: via IS-prosody constraints

A: Kdo nakrmil psa? (Cz)
   ’Who fed the dog?’

B₁: Psa nakrmil Milan. SF *SG NSR
B₂: Milan nakrmil psa. SF *SG NSR
B₃: Milan nakrmil psa. SF *SG NSR
   ‘Milan fed the dog.’

Rationale: Slavic uses free word order to maximize the satisfaction of IS-prosody constraints.

The prosodic hypothesis

Using the constraints assumed for English: STRESS FOCUS (SF), *STRESS GIVEN (*SG), NUCLEAR STRESS RULE (NSR)

Deriving the word order indirectly: via IS-prosody constraints

A: Kdo nakrmil psa? (Cz)
   ‘Who fed the dog?’

B1: Psa nakrmil Milan.

B2: Milan nakrmil psa.

B3: Milan nakrmil psa.
   ‘Milan fed the dog.’

Rationale: Slavic uses free word order to maximize the satisfaction of IS-prosody constraints.

We designed experiments where we manipulate three factors to create patterns predicted by the syntactic and/or the prosodic hypothesis:

- Word order
- Position of stress
- IS-status (focused, given, new/broad focus)
Testing the hypotheses: Modeling part

In order to evaluate the predictions and to estimate the effect sizes of individual constraints in a precise manner, we use Linear Optimality Theory/LOT (Keller 2000; see also Featherston 2005).

In this framework, each constraint is associated with a numeric weight representing the reduction in acceptability caused by a violation of this constraint. If more than one constraint is violated in a sentence, the weights add up in a cumulative manner.

**Procedure:** Each experimental item is manually annotated wrt the constraints that are violated in it. The optimal weight for each constraint based on all experiments is then computed by multiple regression. A separate computation is conducted for the prosodic, syntactic, and combined approach and for each language.
Constraints: The prosodic hypothesis

*STRESS GIVEN
Do not place sentence stress on given constituents.

NUCLEAR STRESS RULE
Place sentence stress on the rightmost constituent in a clause.

Note: We’re leaving STRESS FOCUS aside for the moment.

Constraints: The syntactic hypothesis

*NEW > GIVEN
Do not place a given constituent after a new constituent within a clause.

NUCLEAR STRESS RULE
Place sentence stress on the rightmost constituent in a clause.

Constraints: The unification of both hypotheses

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*STRESS GIVEN</td>
<td>Do not place sentence stress on given constituents.</td>
</tr>
<tr>
<td>*NEW &gt; GIVEN</td>
<td>Do not place a given constituent after a new constituent within a clause.</td>
</tr>
<tr>
<td>NUCLEAR STRESS RULE</td>
<td>Place sentence stress on the rightmost constituent in a clause.</td>
</tr>
</tbody>
</table>

**Note:** The hypotheses are not mutually exclusive.
Experiments: general information

- **Participants**: 40 students in Prague (native speakers of Czech), 40 students in Poznań (native speakers of Polish); 40 students in Bratislava (native speakers of Slovak)

- **Materials**: auditive stimuli consisting of a context utterance and a target sentence; all experiments within a single experimental set-up (forming fillers to each other); always nearly twice as many fillers than experimental items

- **Task/procedure**: participants were instructed to rate the target sentence in the given context; acceptability ratings on a 1–9 scale via computer keyboard. We report normalized z-score values.

**Note**: We only report results of the Czech and Polish experiments. Slovak is very close to Czech. The results are reported in an appendix.
<table>
<thead>
<tr>
<th>Experiment 1</th>
<th>What happens in an all-new context?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 2</td>
<td>Which positions are acceptable for a given object?</td>
</tr>
<tr>
<td>Experiment 3a</td>
<td>Is stress-shift from a given object an alternative to scrambling?</td>
</tr>
<tr>
<td>Experiment 3b</td>
<td>How do focused objects behave?</td>
</tr>
</tbody>
</table>
Experiment 1

factor: position of the object; sentence stress was always on the rightmost element; 32 items (plus 104 fillers)

<table>
<thead>
<tr>
<th>Scrambling the object in an all-new context</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C) Píšou něco zajímavého v novinách?</td>
</tr>
<tr>
<td>‘Is there anything interesting in the newspaper?’</td>
</tr>
<tr>
<td>(a) Dnes prý své zastupitele Němci volí do parlamentu.</td>
</tr>
<tr>
<td>‘Today the Germans allegedly vote their representatives to the parliament.’</td>
</tr>
<tr>
<td>(b) Dnes prý Němci své zastupitele volí do parlamentu.</td>
</tr>
<tr>
<td>(c) Dnes prý Němci volí své zastupitele do parlamentu.</td>
</tr>
<tr>
<td>(d) Dnes prý Němci volí do parlamentu své zastupitele.</td>
</tr>
</tbody>
</table>
In both Czech and Polish, the postverbal conditions are significantly better than the preverbal conditions.
Experiment 2

factor 1: position of the object; factor 2: givenness of the subject; sentence stress was always on the rightmost element; 32 items (plus 104 fillers)

Scrambling a given object (subject is new)

(C) Zjistil jsi, proč dnes sekretářka tak nadávala?
‘Did you find out why our secretary was so angry today?’

(a) Protože prý sekretářku Karel poslal do obchodu.
‘Because Karel allegedly sent the secretary to the store.’

(b) Protože prý Karel sekretářku poslal do obchodu.

(c) Protože prý Karel poslal sekretářku do obchodu.

(d) Protože prý Karel poslal do obchodu sekretářku.
factor 1: position of the object; factor 2: givenness of the subject; sentence stress was always on the rightmost element; 32 items (plus 104 fillers)

### Scrambling a given object (subject is given)

(C) Zjistil jsi, proč dnes sekretářka nadávala na Karla?
 DID you find out why our secretary was so angry with Karel today?

(a) Protože prý sekretářku Karel poslal do obchodu.
 *Because Karel allegedly sent the secretary to the store.*

(b) Protože prý Karel sekretářku poslal do obchodu.

(c) Protože prý Karel poslal sekretářku do obchodu.
 *N > G  SVOPP

(d) Protože prý Karel poslal do obchodu sekretářku.
 *SG *N > G  SVPP0
In both Czech and Polish, the condition where the given object is rightmost/carry stress is judged the worst. In Polish, in addition, there is a preference for a postverbal position of the given object.
Prosodic model (exp 2)

Czech

Polish

z-scores

new subject
given subject

*SG
Syntactic model (exp 2)

Czech

Polish

* N > G  * N > G  * N > G
Combined model (exp 2)

Czech

Polish
factor 1: word order (SVO vs. SOV); factor 2: sentence stress (on V vs. on O); the object was always given; 48 items (plus 88 fillers)

**Stress shift and scrambling (given object)**

(C) Doufám, že ta bouřka nerozbije to okno.
‘I hope that the storm will not break this window.’

(a) Myslím, že učitelka to okno zavřela.
‘I think that the teacher closed this window.’

(b) Myslím, že učitelka zavřela to okno.

(c) Myslím, že učitelka zavřela to okno.

(d) Myslím, že učitelka to okno zavřela.
Results of experiment 3a

In both Czech and Polish, the conditions where the given object is stressed are judged the worst. In addition, Czech prefers scrambling, while Polish prefers stress shift/basic word order.
Prosodic model (exp 3a)

Czech

Polish
Syntactic model (exp 3a)

Czech

Polish

\[ z\text{-scores} \]

\*N > G

NSR

\*N > G

NSR

\*N > G

NSR

\*N > G

NSR

\*N > G

NSR

\*N > G

NSR
Combined model (exp 3a)

Czech

Polish
factor 1: word order (SVO vs. SOV); factor 2: sentence stress (on V vs. on O); the object was always focused; 48 items (plus 88 fillers)

Stress shift and scrambling (focused object)

(C) Nevíš, co učitelka zavřela?
‘Do you know what the teacher closed?’

(a) Myslím, že učitelka to okno zavřela.
‘I think that the teacher closed this window.’

(b) Myslím, že učitelka zavřela to okno. *SG NSR SVO

(c) Myslím, že učitelka zavřela to okno. SVO

(d) Myslím, že učitelka to okno zavřela. NSR *N→G SOV
Results of experiment 3b

In both Czech and Polish, base order and stress on the focused object is the best rated condition. In Polish, scrambling the focused object is relatively acceptable compared to the conditions where the object is not stressed.
Prosodic model (exp 3b)

Czech

Polish
Syntactic model (exp 3b)

Czech

Polish

z-scores

O focused

*N > G NSR

NSR

*N > G
Combined model (exp 3b)

Czech

Polish
Comparison of the models

### Prosodic model

<table>
<thead>
<tr>
<th></th>
<th>Czech</th>
<th></th>
<th>Slovak</th>
<th></th>
<th>Polish</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>SE</td>
<td>estimate</td>
<td>SE</td>
<td>estimate</td>
<td>SE</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.40</td>
<td>0.02</td>
<td>0.29</td>
<td>0.02</td>
<td>0.32</td>
<td>0.02</td>
</tr>
<tr>
<td>*STRESS-GIVEN</td>
<td>-1.00</td>
<td>0.03</td>
<td>-0.79</td>
<td>0.03</td>
<td>-0.92</td>
<td>0.03</td>
</tr>
<tr>
<td>NUCLEAR STRESS RULE</td>
<td>-0.43</td>
<td>0.03</td>
<td>-0.38</td>
<td>0.03</td>
<td>-0.37</td>
<td>0.03</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.2881</td>
<td></td>
<td>0.1851</td>
<td></td>
<td>0.2332</td>
<td></td>
</tr>
</tbody>
</table>

### Syntactic model

<table>
<thead>
<tr>
<th></th>
<th>Czech</th>
<th></th>
<th>Slovak</th>
<th></th>
<th>Polish</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>SE</td>
<td>estimate</td>
<td>SE</td>
<td>estimate</td>
<td>SE</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.44</td>
<td>0.02</td>
<td>0.30</td>
<td>0.02</td>
<td>0.27</td>
<td>0.02</td>
</tr>
<tr>
<td>*NEW&gt;GIVEN</td>
<td>-0.62</td>
<td>0.03</td>
<td>-0.43</td>
<td>0.03</td>
<td>-0.37</td>
<td>0.03</td>
</tr>
<tr>
<td>NUCLEAR STRESS RULE</td>
<td>-0.50</td>
<td>0.03</td>
<td>-0.46</td>
<td>0.04</td>
<td>-0.50</td>
<td>0.04</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.1822</td>
<td></td>
<td>0.1067</td>
<td></td>
<td>0.0987</td>
<td></td>
</tr>
</tbody>
</table>

### Combined model

<table>
<thead>
<tr>
<th></th>
<th>Czech</th>
<th></th>
<th>Slovak</th>
<th></th>
<th>Polish</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>SE</td>
<td>estimate</td>
<td>SE</td>
<td>estimate</td>
<td>SE</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.52</td>
<td>0.02</td>
<td>0.36</td>
<td>0.02</td>
<td>0.34</td>
<td>0.02</td>
</tr>
<tr>
<td>*STRESS-GIVEN</td>
<td>-0.86</td>
<td>0.03</td>
<td>-0.71</td>
<td>0.03</td>
<td>-0.89</td>
<td>0.03</td>
</tr>
<tr>
<td>NUCLEAR STRESS RULE</td>
<td>-0.35</td>
<td>0.03</td>
<td>-0.34</td>
<td>0.03</td>
<td>-0.35</td>
<td>0.03</td>
</tr>
<tr>
<td>*NEW&gt;GIVEN</td>
<td>-0.35</td>
<td>0.03</td>
<td>-0.21</td>
<td>0.03</td>
<td>-0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.3138</td>
<td></td>
<td>0.1939</td>
<td></td>
<td>0.2345</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** For each model all constraints had a significant influence.
Discussion

For each language, a comparison of R-squared values of the models shows that the prosodic model explains more of the variance in the results than the syntactic model. According to ANOVAs, these differences are significant ($p < 0.001$). The combined model is in turn better than the prosodic model; this difference is also significant in all languages ($p < 0.001$ for Czech and Slovak, $p = 0.003$) for Polish. However, the proportion of variance explained by the combined model increases by less than one percent for Slovak and Polish and by less than three percent for Czech.

**Conclusion:** The expression of IS in West Slavic is primarily prosodic.
Cohen (1988) distinguishes between three categories of effect sizes calculated as a factor of standard deviation, where a value around 0.2 standard deviations is considered a small effect, around 0.5 standard deviations a medium effect, and around 0.8 a large effect. We assume that the estimates for the constraint weights correspond to this measure.

### Constraint effect sizes/weights for the combined model

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Czech</th>
<th>Slovak</th>
<th>Polish</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>*STRESS GIVEN</td>
<td>-0.86</td>
<td>-0.71</td>
<td>-0.89</td>
<td>large</td>
</tr>
<tr>
<td>NUCLEAR STRESS RULE</td>
<td>-0.35</td>
<td>-0.34</td>
<td>-0.35</td>
<td>small-medium</td>
</tr>
<tr>
<td>*NEW&gt;GIVEN</td>
<td>-0.35</td>
<td>-0.21</td>
<td>-0.08</td>
<td>small</td>
</tr>
</tbody>
</table>
Further adjustments to the prosodic model

Can the prosodic model be made better by adding further constraints?

We try the following two one-by-one:

**STRESS FOCUS**
Place sentence stress on the focused constituent.

**MOVE**
Do not move constituents (without motivation).

*MOVE: Reinhart 1995, Grimshaw 1997 (formulated as STAY)*
Prosodic model + STRESS FOCUS

Experiment 3b

Stress shift and scrambling (focused object)

(C) Nevíš, co učitelka zavřela?
‘Do you know what the teacher closed?’

(a) Myslím, že učitelka to okno zavřela.
‘I think that the teacher closed this window.’

(b) Myslím, že učitelka zavřela to okno.

(c) Myslím, že učitelka zavřela to okno.

(d) Myslím, že učitelka to okno zavřela.
Prosodic model + STRESS FOCUS (exp 3b)

Czech

Polish
## Comparison of the models

### Prosodic model

<table>
<thead>
<tr>
<th></th>
<th>Czech</th>
<th>SE</th>
<th>Slovak</th>
<th>SE</th>
<th>Polish</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.40</td>
<td>0.02</td>
<td>0.29</td>
<td>0.02</td>
<td>0.31</td>
<td>0.02</td>
</tr>
<tr>
<td>*STRESS-GIVEN</td>
<td>-1.00</td>
<td>0.03</td>
<td>-0.79</td>
<td>0.03</td>
<td>-0.85</td>
<td>0.04</td>
</tr>
<tr>
<td>NUCLEAR STRESS RULE</td>
<td>-0.43</td>
<td>0.03</td>
<td>-0.38</td>
<td>0.03</td>
<td>-0.35</td>
<td>0.03</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.2881</td>
<td></td>
<td>0.1851</td>
<td></td>
<td>0.2332</td>
<td></td>
</tr>
</tbody>
</table>

### Prosodic model + STRESS FOCUS

<table>
<thead>
<tr>
<th></th>
<th>Czech</th>
<th>SE</th>
<th>Slovak</th>
<th>SE</th>
<th>Polish</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.40</td>
<td>0.02</td>
<td>0.29</td>
<td>0.02</td>
<td>0.31</td>
<td>0.02</td>
</tr>
<tr>
<td>*STRESS-GIVEN</td>
<td>-1.03</td>
<td>0.04</td>
<td>-0.79</td>
<td>0.04</td>
<td>-0.85</td>
<td>0.04</td>
</tr>
<tr>
<td>NUCLEAR STRESS RULE</td>
<td>-0.44</td>
<td>0.03</td>
<td>-0.38</td>
<td>0.03</td>
<td>-0.35</td>
<td>0.03</td>
</tr>
<tr>
<td>STRESS FOCUS</td>
<td>0.07</td>
<td>0.05</td>
<td>0.01</td>
<td>0.05</td>
<td>-0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.2883</td>
<td></td>
<td>0.1849</td>
<td></td>
<td>0.2344</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Significant improvement only for Polish (though only by 0.1%).
Discussion

Our results suggest that *STRESS GIVEN is much stronger than STRESS FOCUS or in fact that STRESS FOCUS has no effect at all. One could take this as support for the assumption that only givenness but not focus is a grammatical category (in a broader sense of the word), see e.g. Kadmon & Sevi 2011. But there are reasons to be careful. First, we know that if something is given and focused at the same time, it must receive stress (suggesting that SF is stronger than *SG). Second, our items included plenty of *SG violations but comparatively few SF violations. Moreover, the SF violations present in the items can also be explained by *SG violations (they overlap).

**Conclusion:** Our experiments are not fit to measure the strength of the STRESS FOCUS constraint.
Prosodic model + *MOVE

Experiment 1

Scrambling the object in an all-new context

(C) Píšou něco zajímavého v novinách?  
‘Is there anything interesting in the newspaper?’

(a) Dnes prý své zastupitele Němci volí do parlamentu.  
‘Today the Germans allegedly vote their representatives to the parliament.’

(b) Dnes prý Němci své zastupitele volí do parlamentu.  

(c) Dnes prý Němci volí své zastupitele do parlamentu.

(d) Dnes prý Němci volí do parlamentu své zastupitele.
Prosodic model (exp 1)

Czech

Polish
Prosodic model + *MOVE (exp 1)

Czech

Polish
Prosodic model + *MOVE

Experiment 2

Scrambling a given object (subject is new)

(C) Zjistil jsi, proč dnes sekretářka tak nadávala? ‘Did you find out why our secretary was so angry today?’

(a) Protože prý sekretářku Karel poslal do obchodu. ‘Because Karel allegedly sent the secretary to the store.’

(b) Protože prý Karel sekretářku poslal do obchodu. *MOVE SOVPP

(c) Protože prý Karel poslal sekretářku do obchodu. SVOPP

(d) Protože prý Karel poslal do obchodu sekretářku. *SG SVPOPO
Prosodic model (exp 2)

Czech

Polish
Prosodic model + *MOVE (exp 2)

Czech

Polish

-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
z-scores
new subject
given subject
O S V PP S O V PP S V O PP S V PP O
z-scores
*MOVE *MOVE
*SG

-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
z-scores
new subject
given subject
O S V PP S O V PP S V O PP S V PP O
z-scores
*MOVE *MOVE
*SG
Experiment 3a

Stress shift and scrambling (given object)

(C) Doufám, že ta bouřka nerozbije to okno.
‘I hope that the storm will not break this window.’

(a) Myslím, že učitelka to okno zavřela.
‘I think that the teacher closed this window.’

(b) Myslím, že učitelka zavřela to okno.

(c) Myslím, že učitelka zavřela to okno.

(d) Myslím, že učitelka to okno zavřela.
Prosodic model (exp 3a)

Czech

Polish

![Graphs showing z-scores for Czech and Polish languages with labels for NSR and *SG categories.](image-url)
Prosodic model + *MOVE (exp 3a)

Czech

Polish

![Graphs comparing z-scores for Czech and Polish languages with labels for *MOVE, NSR, and *SG.](image-url)
Prosodic model + *MOVE

Experiment 3b

Stress shift and scrambling (focused object)

(C) Nevíš, co učitelka zavřela?
‘Do you know what the teacher closed?’

(a) Myslím, že učitelka to okno zavřela.
‘I think that the teacher closed this window.’

(b) Myslím, že učitelka zavřela to okno.

(c) Myslím, že učitelka zavřela to okno.

(d) Myslím, že učitelka to okno zavřela.
Prosodic model (exp 3b)

Czech

Polish
Prosodic model + *MOVE (exp 3b)

Czech

Polish
Comparison of the models

### Prosodic model

<table>
<thead>
<tr>
<th></th>
<th>Czech</th>
<th>Slovak</th>
<th>Polish</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(Intercept)</strong></td>
<td>0.40</td>
<td>0.29</td>
<td>0.32</td>
</tr>
<tr>
<td>*STRESS-GIVEN</td>
<td>-1.00</td>
<td>-0.79</td>
<td>-0.92</td>
</tr>
<tr>
<td>NUCLEAR STRESS RULE</td>
<td>-0.43</td>
<td>-0.38</td>
<td>-0.37</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td><strong>0.2881</strong></td>
<td><strong>0.1851</strong></td>
<td><strong>0.2332</strong></td>
</tr>
</tbody>
</table>

### Prosodic model + *MOVE

<table>
<thead>
<tr>
<th></th>
<th>Czech</th>
<th>Slovak</th>
<th>Polish</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(Intercept)</strong></td>
<td>0.51</td>
<td>0.46</td>
<td>0.64</td>
</tr>
<tr>
<td>*STRESS-GIVEN</td>
<td>-1.04</td>
<td>-0.85</td>
<td>-1.03</td>
</tr>
<tr>
<td>NUCLEAR STRESS RULE</td>
<td>-0.42</td>
<td>-0.36</td>
<td>-0.34</td>
</tr>
<tr>
<td>*MOVE</td>
<td>0.21</td>
<td>0.30</td>
<td>-0.59</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td><strong>0.2997</strong></td>
<td><strong>0.2083</strong></td>
<td><strong>0.3208</strong></td>
</tr>
</tbody>
</table>

**Note:** Significant improvement for all three languages, though most clear for Polish (8.7%).
Discussion

Our results show that the *MOVE constraint (Reinhart, Grimshaw) might be operative in all three languages. For Czech and Slovak, its effect is rather small. This is due to the conflicting effects observed for focused and given constituents: while focused constituents clearly disprefer scrambling (exp 3b), given constituents prefer scrambling over stress shift (exp3a) and freely allow scrambling without any effect on the output (exp 2). The situation is different in Polish, where scrambling is the dispreferred option independently of the IS-status of the constituent (exp 2, 3a, 3b), yielding the comparatively large effect of the *MOVE constraint in the model as well as the overall improvement of the model with *MOVE.

Conclusion: Parametrization within West Slavic wrt to whether (unmotivated) movement is penalized.
Overall conclusion

How is information structure expressed in West Slavic languages?
Overall conclusion

How is information structure expressed in West Slavic languages?

Primarily prosodically, secondarily by word order. More particularly:
How is information structure expressed in West Slavic languages?

Primarily prosodically, secondarily by word order. More particularly:

- Most variation in our data is explained by prosodic constraints (*STRESS GIVEN and NUCLEAR STRESS RULE).
How is information structure expressed in West Slavic languages?

Primarily prosodically, secondarily by word order. More particularly:

- Most variation in our data is explained by prosodic constraints (*STRESS GIVEN and NUCLEAR STRESS RULE).
- Adding a word-order constraint (*NEW> GIVEN) results only in a slight improvement of the model.
Overall conclusion

How is information structure expressed in West Slavic languages?

Primarily prosodically, secondarily by word order. More particularly:

- Most variation in our data is explained by prosodic constraints (*STRESS GIVEN and NUCLEAR STRESS RULE).
- Adding a word-order constraint (*NEW > GIVEN) results only in a slight improvement of the model.

The role of other constraints:

- There is a strong influence of *MOVE for Polish.
- The role of STRESS FOCUS is an open issue.
We interpret our overall results as an argument in favor of “autonomous syntax”, i.e. a syntactic system that is “uncontaminated by information structure” and potentially other pragmatic features (Fanselow 2006, 2008, Horváth 2010, Fanselow & Lenertová 2011). We see that even in discourse-configurational languages like the West Slavic languages, prosody is primary in the expression of information structure. Syntax is involved only secondarily, as a tool to maximize constraint satisfaction. The effect of purely syntactic constraints on IS (*NEW>GIVEN) is found to be minor and it is a matter of future research to determine whether it cannot be attributed to non-IS factors (definiteness/specificity/presuppositionality).
Outlook

- Balance constraint violation across the experimental setup.
- Discern scrambling from other types of movement (A from A′ movement; ad exp 1, 2: the OSVPP condition; ad exp 3b: compare focus scrambling to focus A′ movement).
- Discern givenness from definiteness/specificity (cf. Šimík & Wierzba, under review).
- If possible, test further syntactic hypotheses.


References

References

Appendix

Results and models for Slovak (as compared to Czech)
Results of experiment 1

Slovak

Czech
Prosodic model (exp 1)

Slovak

Czech
Prosodic model + *MOVE (exp 1)

Slovak

Czech

-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
all new
O S V PP S O V PP S V O PP S V PP O
*MOVE*MOVE *MOVE*MOVE

-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
all new
O S V PP S O V PP S V O PP S V PP O
*MOVE*MOVE *MOVE*MOVE
Results of experiment 2

Slovak

Czech

-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
z-scores
new subject
given subject
O S V PP S O V PP S V O PP S V PP O

-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
z-scores
new subject
given subject
O S V PP S O V PP S V O PP S V PP O
Prosodic model (exp 2)

Slovak

Czech

*SG
Syntactic model (exp 2)

Slovak

Czech

* N > G * N > G * N > G

* N > G * N > G * N > G
Combined model (exp 2)

Slovak

Czech
Prosodic model + *MOVE (exp 2)

Slovak

Czech

![Graphs showing z-scores for Slovak and Czech languages with new subject and given subject marked by circles and dots respectively.]
Results of experiment 3a

Slovak

Czech
Prosodic model (exp 3a)

Slovak

\[
\begin{array}{cccccccc}
O & V & V & O & V & O & O & V \\
\end{array}
\]

Czech

\[
\begin{array}{cccccccc}
O & V & V & O & V & O & O & V \\
\end{array}
\]
Syntactic model (exp 3a)

Slovak

Czech

*z > G
NSR
*z > G
NSR
*z > G
NSR
*z > G
NSR
Combined model (exp 3a)

### Slovak

![Graph showing z-scores for Slovak model with annotations *N > G, *NSR, *SG.

### Czech

![Graph showing z-scores for Czech model with annotations *N > G, *NSR, *SG.}
Prosodic model + *MOVE (exp 3a)

Slovak

Czech

\begin{align*}
\text{O given} & \quad \text{O} \quad \text{V} \quad \text{O} \quad \text{V} \quad \text{O} \quad \text{V} \\
*\text{MOVE} & \quad \text{NSR} \quad *\text{SG}
\end{align*}
Results of experiment 3b

**Slovak**

<table>
<thead>
<tr>
<th>z-scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.5</td>
</tr>
<tr>
<td>-1.0</td>
</tr>
<tr>
<td>-0.5</td>
</tr>
<tr>
<td>0.0</td>
</tr>
<tr>
<td>0.5</td>
</tr>
<tr>
<td>1.0</td>
</tr>
<tr>
<td>1.5</td>
</tr>
</tbody>
</table>

**Czech**

<table>
<thead>
<tr>
<th>z-scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.5</td>
</tr>
<tr>
<td>-1.0</td>
</tr>
<tr>
<td>-0.5</td>
</tr>
<tr>
<td>0.0</td>
</tr>
<tr>
<td>0.5</td>
</tr>
<tr>
<td>1.0</td>
</tr>
<tr>
<td>1.5</td>
</tr>
</tbody>
</table>
Prosodic model (exp 3b)

Slovak

Czech
Syntactic model (exp 3b)

Slovak

Czech

\[ z \text{-scores} \]

\[ * N > G \quad \text{NSR} \]

\[ * N > G \quad \text{NSR} \]
Combined model (exp 3b)

Slovak

Czech
Prosodic model + STRESS FOCUS (exp 3b)

Slovak

Czech

-1.5 −1.0 −0.5 0.0 0.5 1.0 1.5
z−scores
O focused
O V V O V O O V
*SG
SF
*SG
NSR
SF
NSR
Prosodic model + *MOVE (exp 3b)

Slovak

Czech

![Graphs comparing Slovak and Czech prosodic models with *MOVE](image-url)