Evaluation of frame-semantic role labeling in a case-marking language

Olga Lyashevskaya\textsuperscript{1,2} (olesar@gmail.com)
Egor Kashkin\textsuperscript{2} (egorkashkin@rambler.ru)
\textsuperscript{1}Higher School of Economics, Moscow
\textsuperscript{2}Vinogradov Institute of Russian Language RAS, Moscow
SEMANTIC ROLE LABELING (SRL): BACKGROUND

'History does not take into account privileges'

'[She/He] does not know history'
SEMANTIC ROLE LABELING (SRL): BACKGROUND

'History does not take into account privileges'

'[She/He] does not know history'

CoNLL 2008 shared task (Surdeanu et al. 2008): PropBank-like SRL
SemEval 2007 (Baker et al. 2007): FrameNet-like SRL
SRL FOR RUSSIAN

- Anisimovich et al. 2012, Petrova 2013
- Azarova 2008
- Shelmanov & Smirnov 2014
- Ermakov & Pleshko 2009
- Kuznetsov 2012

... yet project in progress

- Many possible classifications of roles and frames

- Russian FrameBank as a benchmark for SRL evaluation + class mapping
OUTLINE OF EVALUATION

EXPERIMENT

- a prototype of SRL module (rule-based)
- training and test data sets from FrameBank
- how it works (cues and challenges of Russian PP SRL)
- induced roles VS gold standard roles:
  - metrics
  - the goodness of fit for non-matching pairs
DATA SOURCE: RUSSIAN FRAMEBANK

- Berkeley FrameNet: extralinguistic situations (frames) → a set of participants → lexemes.
- Russian FrameBank (www.framebank.ru): lexemes (about 2200, primarily verbs) → constructions ([Apresjan, Pall 1982] & added by annotators) → examples from RNC (about 100 for each lexeme, manually tagged)
- Theoretical framework: FrameNet and Construction Grammar + Moscow Semantic School

резать ‘to cut, to carve’:
✓ Продавщица режет сыр ‘The shop assistant is cutting cheese’; Она режет хлеб на тонкие куски ‘She is slicing bread (lit.: cutting bread into thin slices)’…
✓ Старик резал четки из кипариса ‘The old man carved rosaries out of cypress’; Он резал деревянные ложки ножом ‘He carved wooden spoons with a knife’…
✓ У него в желудке резало ‘He had griping pains in his stomach (lit.: It was cutting in his stomach)’
**FRAMEBANK: construction pattern**

<table>
<thead>
<tr>
<th>ID230</th>
<th>Cx name: <em>Piatno vystupilo na rubaške</em> ['a stain appeared on the short']. Cx Pattern: Snom V na + Sloc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cx Item ID</td>
<td>Pl</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>2077</td>
<td>1</td>
</tr>
<tr>
<td>2079</td>
<td>3</td>
</tr>
</tbody>
</table>

Lexical Index of target words

Index of Morphosyntactic Items
# FrameBank: Tagging

<table>
<thead>
<tr>
<th>ID</th>
<th>Pl</th>
<th>Letter</th>
<th>Head</th>
<th>Phrase</th>
<th>Explication</th>
<th>Syntactic Rank</th>
<th>Lexico-semantic constraints</th>
<th>Status / Realization</th>
</tr>
</thead>
<tbody>
<tr>
<td>18589</td>
<td>1</td>
<td>X</td>
<td>Snom</td>
<td>NPnom</td>
<td>Agent</td>
<td>Subject</td>
<td>human</td>
<td>Oblig.</td>
</tr>
<tr>
<td>20089</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>No</td>
<td>-</td>
<td>Omitted. Licensed by Imperative Cx</td>
</tr>
<tr>
<td>18590</td>
<td>2</td>
<td>-</td>
<td>sobrat’</td>
<td>-</td>
<td>to collect</td>
<td>Predicate</td>
<td>-</td>
<td>Oblig.</td>
</tr>
<tr>
<td>20090</td>
<td>1</td>
<td>-</td>
<td>soberi.Vimper</td>
<td>-</td>
<td>to collect</td>
<td>Predicate</td>
<td>-</td>
<td>Standard</td>
</tr>
<tr>
<td>18591</td>
<td>3</td>
<td>Y</td>
<td>Vinf</td>
<td>VPinf</td>
<td>what is collected</td>
<td>Peripheral</td>
<td>eat</td>
<td>Oblig.</td>
</tr>
<tr>
<td>20091</td>
<td>3</td>
<td>Y</td>
<td>poest’.Vinf</td>
<td>poest’ v dorogu.VPinf</td>
<td>what is collected</td>
<td>Peripheral</td>
<td>eat</td>
<td>Standard</td>
</tr>
<tr>
<td>20089</td>
<td>2</td>
<td>Z</td>
<td>nam.SPROdat</td>
<td>nam.NPdat</td>
<td>Beneficiary</td>
<td>Peripheral</td>
<td>human</td>
<td>Added. Licensed by Ditransitive Dative Cx</td>
</tr>
<tr>
<td>20089</td>
<td>4</td>
<td>W</td>
<td>v dorogu</td>
<td>v dorogu.NPdat</td>
<td>Goal</td>
<td>Adjunct</td>
<td>abstract</td>
<td>Added</td>
</tr>
</tbody>
</table>
FrameBank: Semantic Roles

- Different inventories of semantic roles (cf. Ch. Fillmore, Ju.D. Apresjan, E.V. Paducheva, etc.)

- FrameBank:
  - hierarchy of semantic roles → flexible search options
  - correlation between the roles and the semantic classes of verbs.
  - semantic roles graph: 96 items → 6 domains → further smaller units.
FRAME BANK: SEMANTIC ROLES GRAPH

Subject_of_mental_state [Cognizer]

Subject_of_perception

Subject_of_physiological_state

Subject_of_physiological_response

Experiencer

Beneficiary

Addresssee

Audience

Limit
Result

Patient_of_motion

Patient_of_social_relation

Content [Mental_content]

Message

Topic_of_thought

Topic

Part_of_patient_in_focus

Part_of_theme_in_focus

Stimulus

Smell

Event_in_focus [Event]
OUTLINE OF EVALUATION
EXPERIMENT

- a prototype of SRL module (rule-based)
- training and test data sets from FrameBank
- how it works (cues and challenges of Russian PP SRL)
- induced roles VS gold standard roles:
  - metrics
  - the goodness of fit for non-matching pairs
SRL PROTOTYPE: PREPOSITIONAL

PPs

- Four PPs:
  - za + NPins
  - za + NPacc
  - ot + NPgen
  - po + NPdat

- Very frequent (e.g. ca. 900,000 hits of “ot + NPgen” in the RNC).

- Highly polysemous (e.g. 14 possible roles of “za + NPacc”)

TRAINING AND TEST DATA

- Training data set: constructions from [Apresjan, Pall 1982]
- Test data set: constructions added by annotators

- NB type units (from dictionary), not tokens (hits from corpus)
<table>
<thead>
<tr>
<th>PP</th>
<th>Training set: ‘old’ data</th>
<th>Test set: ‘new’ data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># constructions</td>
<td># examples</td>
</tr>
<tr>
<td>za + NPins</td>
<td>95</td>
<td>80</td>
</tr>
<tr>
<td>za + NPacc</td>
<td>228</td>
<td>223</td>
</tr>
<tr>
<td>ot + NPgen</td>
<td>266</td>
<td>435</td>
</tr>
<tr>
<td>po + NPdat</td>
<td>311</td>
<td>245</td>
</tr>
<tr>
<td>Total</td>
<td>900</td>
<td>983</td>
</tr>
</tbody>
</table>
SRL PROTOTYPE: PREPOSITIONAL PPs

- 62 heuristics

  - morphosyntactic pattern
  - lexical class of the PP argument
  - lexical class of the predicate
  - lexical classes of other arguments

Examples: za + NPacc

- NP\text{nom} V_{\text{CHANGE POSS.}}. \text{NPacc} za + \text{NPacc} \rightarrow \text{Price}
- za + \text{NPacc}_{\text{TIME PERIOD}} \rightarrow \text{Period}

cf. [Toldova et al. 2008] on the semantic filters in the RNC
SRL PROTOTYPE: PREPOSITIONAL PPs

- Two possible outcomes for some rules:
  \[ \text{NPnom}_{\text{ANIM}} \ V_{\text{MOTION}} \ za + \ \text{NPins}_{\text{ANIM}} \]

- Counter-Agent: *Milicioner pobezhal za prestupnikom* ‘A policeman ran after an offender’
- Goal: *Mal’chik pobezhal v bol’nicu za vrachom* ‘A boy ran to hospital to call the doctor’
SRL CUES

- Semantic class of a verb
  - *Beshus’*_{EMOTION} za doch moju ‘I am in a rage because of my daughter’ → Reason for an Emotion, cf. bojat’sja ‘to be afraid’, bespokoit’sja ‘to worry about sth.’

- Semantic class of a participant
  - *Po radio*{COMMUNIC. FACIL.} igrala muzyka ‘There was music broadcast (lit.: played) by radio’ → Manner. cf. vystupat’ *po televizoru* ‘to speak on TV’, poslat’ dokumenty *po pochте* ‘to send documents by post’
CHALLENGES FOR SRL

- Pragmatic factors, cf. Counter-Agent vs. Goal of motion.
- No clear semantic constraints on a verb or on its arguments.
  - “po + NPdat???” for Reason: Rasskaz byl prochitan po ego pros’be ‘The story was read at his request’ → Information Source, Reason, cf. zhenit’sja po ljubvi ‘to make a love-match (lit.: to get married due to love)’, uvolit’ po sokrascheniju shtatov ‘to discharge sb. on grounds of staff reduction’, sidet’ zdes’ po drugomu delu ‘to stay here on some other business’
- Metonymies of concrete nouns
  - Ego nevozmozhno otorvat’ ot knigi ‘It is impossible to divert his attention from the book (lit.: to tear him=it from the book)’ → Patient & Location Content of Action
EVALUATION METRICS

1. Precision, Recall, F1

\[ P = \frac{\text{#Matches}}{\text{#E_Answers}} \]
\[ R = \frac{\text{#Matches}}{\text{#G_Answers}} \]

2. Cluster purity and collocation, F1 (Lang & Lapata 2011)

\[ P_u = \frac{1}{n} \sum_{i=1}^{n_C} \max_{j=1,\ldots,n_C} |C_i \cap G_j| \]
\[ C_o = \frac{1}{n} \sum_{j=1}^{n_C} \max_{i=1,\ldots,n_C} |C_i \cap G_j| \]

3. Repulsion: how likely are the roles to distinguish frames

\[ \text{repulsion} = \frac{\#\text{Verbs(RoleE\_RoleG OR RoleE!RoleG)}}{\sqrt{\#\text{Verbs(RoleE)} \times \#\text{Verbs(RoleG)}}} \]

4. Distance between roles in semantic role graph
# RESULTS: PRECISION

<table>
<thead>
<tr>
<th>PP</th>
<th>Total amount of new patterns</th>
<th>‘Strong’ matching (the role is identified correctly and unambiguously)</th>
<th>‘Weak’ matching (one of the answers is correct)</th>
<th>$P_{\text{strong}}$</th>
<th>$P_{\text{strong+weak}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>za + NPins</td>
<td>19</td>
<td>9</td>
<td>7</td>
<td>0.47</td>
<td>0.84</td>
</tr>
<tr>
<td>za + NPacc</td>
<td>37</td>
<td>22</td>
<td>11</td>
<td>0.59</td>
<td>0.89</td>
</tr>
<tr>
<td>ot + NPgen</td>
<td>70</td>
<td>41</td>
<td>24</td>
<td>0.59</td>
<td>0.93</td>
</tr>
<tr>
<td>po + NPdat</td>
<td>65</td>
<td>32</td>
<td>25</td>
<td>0.49</td>
<td>0.88</td>
</tr>
<tr>
<td>Total</td>
<td>191</td>
<td>104</td>
<td>67</td>
<td>0.54</td>
<td>0.90</td>
</tr>
</tbody>
</table>
THE GOODNESS OF FIT AND REPULSION

<table>
<thead>
<tr>
<th>Matching Evaluation (human)</th>
<th>Role E</th>
<th>Role G</th>
<th>#Verbs (RoleE)</th>
<th>#Verbs (RoleG)</th>
<th>#Verbs (RoleE! RoleG)</th>
<th>#Verbs (RoleG! RoleE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Source</td>
<td>Reason</td>
<td>12</td>
<td>266</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

*Lovit' kaif [ot knig] ‘To be in high from books’.*

| Good                        | Path   | Patient | 105           | 712           | 46                   |

*On bedit i mechetsja golovoj [po perekladine] ‘He raves, tossing his head over the crossbar’.*

| Good                        | Property | Reason | 175           | 266           | 31                   |

*Ego zabrali [po nacional'nomu piznaku] ‘He was arrested on ethnic grounds’.*

| Average                     | Term    | Time_point | 52           | 42            | 6                    |

*Vstrecha prodlilas' [za polnoch] ‘The meeting lasted past midnight’.*

| Average                     | Target_ |                      |              |               |                      |
### THE GOODNESS OF FIT AND REPULSION

<table>
<thead>
<tr>
<th>Role</th>
<th>#Verbs (RoleE)</th>
<th>#Verbs (RoleG)</th>
<th>#Verbs (RoleE! RoleG)</th>
<th>#Verbs (RoleE+ RoleG)</th>
<th>Repulsion</th>
<th>Same domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reason</td>
<td>12</td>
<td>266</td>
<td>3</td>
<td>0</td>
<td>0.05</td>
<td>NO</td>
</tr>
</tbody>
</table>

- **Reason**
  - [Source↑External_cause↑Agent]↑Root↓[Setting↓Reason]
  - 'The voters take a vote through the lists'.

<table>
<thead>
<tr>
<th>Role</th>
<th>#Verbs (RoleE)</th>
<th>#Verbs (RoleG)</th>
<th>#Verbs (RoleE! RoleG)</th>
<th>#Verbs (RoleE+ RoleG)</th>
<th>Repulsion</th>
<th>Same domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td>105</td>
<td>712</td>
<td>46</td>
<td>3</td>
<td>0.18</td>
<td>NO</td>
</tr>
</tbody>
</table>

- **Patient**
  - [Path↑Location↑Setting]↑Root↓[Patient]
  - 'He raves, tossing his head over the crossbar'.

<table>
<thead>
<tr>
<th>Role</th>
<th>#Verbs (RoleE)</th>
<th>#Verbs (RoleG)</th>
<th>#Verbs (RoleE! RoleG)</th>
<th>#Verbs (RoleE+ RoleG)</th>
<th>Repulsion</th>
<th>Same domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reason</td>
<td>175</td>
<td>266</td>
<td>31</td>
<td>5</td>
<td>0.17</td>
<td>(YES)</td>
</tr>
</tbody>
</table>

- **Reason**
  - [Property↑Setting↓Reason]
  - 'He was arrested on ethnic grounds'.

<table>
<thead>
<tr>
<th>Role</th>
<th>#Verbs (RoleE)</th>
<th>#Verbs (RoleG)</th>
<th>#Verbs (RoleE! RoleG)</th>
<th>#Verbs (RoleE+ RoleG)</th>
<th>Repulsion</th>
<th>Same domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time_point</td>
<td>52</td>
<td>42</td>
<td>6</td>
<td>2</td>
<td>0.17</td>
<td>YES</td>
</tr>
</tbody>
</table>

- **Time_point**
  - [Term↓Time_point]
  - 'The meeting lasted past midnight'.

<table>
<thead>
<tr>
<th>Role</th>
<th>#Verbs (RoleE)</th>
<th>#Verbs (RoleG)</th>
<th>#Verbs (RoleE! RoleG)</th>
<th>#Verbs (RoleE+ RoleG)</th>
<th>Repulsion</th>
<th>Same domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target_</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CONCLUSION

- Repulsion correlate quite well with the split between Average and Bad matches (repulsion threshold .20) \( \checkmark \)
- ... but split between Good and Average matches:: FAILED \( \chi \)
- Same-domain VS Good_Average_Bad matches:
  - all Bad matches do not share the same domain but this is not sufficient \( \chi \)
- Graph-based distances: FAILED \( \chi \)
- *Future development*
  - Other SRL modules
  - Token frequencies
  - More (other than IS-A) types of edges in the SR graph
СПАСИБО!