Automatic identification of shared arguments in verbal coordinations

Aleksandrs Berdičevskis, Hanne Eckhoff
Александр Бердичевский, Ханне Экхофф

“Dialogue”, Moscow, April 27, 2015
Example to start with
Example to start with

Петя собирает и ест ягоды
Peter is_picking and is_eating berries

Петя курит и ест ягоды
Peter is_smoking and is_eating berries
Example to start with

Петя собирает и ест ягоды
Peter is_picking and is_eating berries

Петя курит и ест ягоды
Peter is_smoking and is_eating berries

Are the syntactic structures different?
Example to start with

Петя собирает и ест ягоды
Peter is_picking and is_eating berries

Петя курит и ест ягоды
Peter is_smoking and is_eating berries

Yes, arguments are shared differently.
Example to start with

Петя собирает и ест ягоды
Peter is_picking and is_eating berries

Петя курит и ест ягоды
Peter is_smoking and is_eating berries

Yes, arguments are shared differently.
Example to start with

Our goal: restore the information about shared arguments, give only the “primary” syntactic structure.
Example to start with

Петя собирает и ест ягоды
Peter is_picking and is_eating berries

Петя курит и ест ягоды
Peter is_smoking and is_eating berries

Our goal: restore the information about shared arguments, give only the “primary” syntactic structure.
Why do we care?
Outline of the talk

• Why do we care? (Background and problem)
• What do we do? (Materials and methods)
• What do we get? (Results and perspectives)
I. WHY DO WE CARE?
Why do we care?

• “Birds & Beasts” project: diachrony of Russian aspect
Why do we care?

• “Birds & Beasts” project: diachrony of Russian aspect
  – One of the questions: are/were there any differences in the argument-structure preferences of perfective and imperfective verbs?
Why do we care?

• “Birds & Beasts” project: diachrony of Russian aspect
  – One of the questions: are/were there any difference in the argument-structure preferences of perfective and imperfective verbs?
    • Investigation tool: TOROT, a diachronic treebank
      – Old Church Slavonic
      – Old Russian
      – Middle Russian
      – Modern Russian
Why do we care?

- “Birds & Beasts” project: diachrony of Russian aspect
  - One of the questions: are/were there any differences in the argument-structure preferences of perfective and imperfective verbs?
- Investigation tool: TOROT, a diachronic treebank
  - Old Church Slavonic
  - Old Russian
  - Middle Russian
  - Modern Russian

≈250000 words, developed at the Us of Oslo and Tromsø
Why do we care?

• “Birds & Beasts” project: diachrony of Russian aspect
  – One of the questions: are/were there any differences in the argument-structure preferences of perfective and imperfective verbs?
  • Investigation tool: TOROT, a diachronic treebank
    – Old Church Slavonic
    – Old Russian
    – Middle Russian
    – Modern Russian

≈250000 words, developed at the Us of Oslo and Tromsø

= Syntagrus (≈860000 words, developed at the IITP, converted to the TOROT format by us)
Why do we care?

• “Birds & Beasts” project: diachrony of Russian aspect
  – One of the questions: are/were there any difference in the argument-structure preferences of perfective and imperfective verbs?
  
• Investigation tool: TOROT, a diachronic treebank
  – Old Church Slavonic
  – Old Russian
  – Middle Russian
  – Modern Russian

≈250000 words, developed at the Us of Oslo and Tromsø

= Syntagrus (≈860000 words, developed at the IITP, converted to the TOROT format by us)
# The two dependency annotation schemes

<table>
<thead>
<tr>
<th><strong>Syntagrus</strong></th>
<th><strong>TOROT (PROIEL)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>– Based on the “Meaning-Text” Theory, strongly relies on lexical semantics</td>
<td>– Inspired by Lexical-Functional Grammar</td>
</tr>
<tr>
<td>– 60+ relation labels</td>
<td>– 20+ relation labels</td>
</tr>
<tr>
<td>– Limited use of empty tokens</td>
<td>– Active use of empty tokens (verbs and conjunctions)</td>
</tr>
<tr>
<td>– Primary dependencies only</td>
<td>– Secondary dependencies (external subjects; predicate identity in gapping; shared dependents)</td>
</tr>
</tbody>
</table>
The two dependency annotation schemes

<table>
<thead>
<tr>
<th>Syntagrus</th>
<th>TOROT (PROIEL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Based on the “Meaning-Text” Theory, strongly relies on lexical semantics</td>
<td>– Inspired by Lexical-Functional Grammar</td>
</tr>
<tr>
<td>– 60+ relation labels</td>
<td>– 20+ relation labels</td>
</tr>
<tr>
<td>– Limited use of empty tokens</td>
<td>– Active use of empty tokens (verbs and conjunctions)</td>
</tr>
<tr>
<td>– Primary dependencies only</td>
<td>– Secondary dependencies (external subjects; predicate identity in gapping; shared dependents)</td>
</tr>
</tbody>
</table>
The two dependency annotation schemes

<table>
<thead>
<tr>
<th>Syntagrus</th>
<th>TOROT (PROIEL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Based on the “Meaning-Text” Theory, strongly relies on lexical semantics</td>
<td>– Inspired by Lexical-Functional Grammar</td>
</tr>
<tr>
<td>– 60+ relation labels</td>
<td>– 20+ relation labels</td>
</tr>
<tr>
<td>– Limited use of empty tokens</td>
<td>– Active use of empty tokens (verbs and conjunctions)</td>
</tr>
<tr>
<td>– Primary dependencies only</td>
<td>– Secondary dependencies (external subjects; predicate identity in gapping; shared dependents)</td>
</tr>
</tbody>
</table>

Different approach to coordination
Coordination and dependency syntax

• Coordination per se is a problem for dependency frameworks due to its *symmetric* nature

• Popel et al. 2013:
  – One of the most frequent sources of parsing errors
  – Three major annotation styles ("Moscow", "Prague" and "Stanford")
    • ...but virtually every treebank has its own substyle
  – No reliable solution for recovering shared dependents from the primary dependency structure
Coordination representation

Петя собирает ягоды
Peter is_picking and is_eating berries

Syntagrus
Coordination representation

Петя собирает и ест ягоды

Peter is_picking and is_eating berries
Coordination representation

Петя собирает ягоды
Peter is_picking and is_eating berries

Петя ест ягоды
Peter is_eating berries
II. WHAT DO WE DO?
We limit ourselves to verbs and their arguments

\[ \text{sub} = \text{subject (John wrote a book)} \]
\[ \text{obj} = \text{direct object (John wrote a book)} \]
\[ \text{obl} = \text{indirect/oblique object (John gave me a book)} \]
\[ \text{comp} = \text{complement clause (John said that he wrote a book)} \]
\[ \text{xobj} = \text{open embedded predication (John is a writer; John decided to write a book)} \]
\[ \text{ag} = \text{passive agent (The book was written by John)} \]
Identifying datapoints: the empty-slot criterion
Identifying datapoints: the empty-slot criterion

Conjunction

C

PRED

VERB1

PRED

ARGUMENT1

PRED

VERB2

OBJ

ARGUMENT2

OBJ

ARGUMENT3
Identifying datapoints: the empty-slot criterion

Conjunction

Coordinated verbs

ARGUMENT1
ARGUMENT2
ARGUMENT3
Identifying datapoints: the empty-slot criterion

Conjunction

Coordinated verbs

Both have an OBJ, i.e. both OBJ slots are filled
Identifying datapoints: the empty-slot criterion

Verb1 has a SUB, while Verb2 has an empty SUB slot

Both have an OBJ, i.e. both OBJ slots are filled

Conjunction

Coordinated verbs
Identifying datapoints: the empty-slot criterion

Verb1 has a SUB, while Verb2 has an empty SUB slot

Argument1 can be a shared dependent.
{Verb2; Argument1} -- a datapoint; a potential adoption.
Training sample

- We extracted 1103 datapoints (346 sentences) from the converted Syntagrus
- The shared dependencies were manually inserted (≈15 hours of work for an unexperienced annotator)
Training sample

- We extracted 1103 datapoints (346 sentences) from the converted Syntagrus
- The shared dependencies were manually inserted (≈15 hours of work for an unexperienced annotator)

<table>
<thead>
<tr>
<th>Argument type</th>
<th>Adoption rate</th>
<th>Potential adoptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUB</td>
<td>0.80</td>
<td>261</td>
</tr>
<tr>
<td>OBJ</td>
<td>0.06</td>
<td>251</td>
</tr>
<tr>
<td>OBL</td>
<td>0.02</td>
<td>393</td>
</tr>
<tr>
<td>XOBJ</td>
<td>0.01</td>
<td>158</td>
</tr>
<tr>
<td>COMP</td>
<td>0.00</td>
<td>36</td>
</tr>
<tr>
<td>AG</td>
<td>0.00</td>
<td>4</td>
</tr>
</tbody>
</table>
Training sample

- We extracted 1103 datapoints (346 sentences) from the converted Syntagrus
- The shared dependencies were manually inserted (≈15 hours of work for an unexperienced annotator)

<table>
<thead>
<tr>
<th>Argument type</th>
<th>Adoption rate</th>
<th>Potential adoptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUB</td>
<td>0.80</td>
<td>261</td>
</tr>
<tr>
<td>OBJ</td>
<td>0.06</td>
<td>251</td>
</tr>
<tr>
<td>OBL</td>
<td>0.02</td>
<td>393</td>
</tr>
<tr>
<td>XOBJ</td>
<td>0.01</td>
<td>158</td>
</tr>
<tr>
<td>COMP</td>
<td>0.00</td>
<td>36</td>
</tr>
<tr>
<td>AG</td>
<td>0.00</td>
<td>4</td>
</tr>
</tbody>
</table>
Training the guesser: probabilistic features

Feature 1: the probability of a potential adopter (specific verb) having an argument of the adoptee type (SUB, OBJ...).
Training the guesser: probabilistic features

Feature 1: the probability of a potential adopter (specific verb) having an argument of the adoptee type (SUB, OBJ...)

Петя курит и ест ягоды
Peter is_smoking and is_eating berries

Datapoint 1: {курить; OBJ} {is smoking; berries (OBJ)}
Feature 1: 0.13
Datapoint 2: {есть; SUB} {is_eating; Peter (SUB)}
Feature 1: 0.30
Training the guesser: probabilistic features

Feature 2: the probability of a potential adopter having an argument frame that would consist of its own primary arguments and a potential adoptee.

Петя курит и ест ягоды
Peter is_smoking and is_eating berries

Datapoint 1: {курить; SUB+OBJ} {is_smoking; berries (OBJ)}
Feature 2: 0.00

Datapoint 2: {есть; SUB+OBJ} {is_eating; Peter (SUB)}
Feature 2: 0.18
Training the guesser: probabilistic features

Feature 3: the probability of a particular argument being an adoptee

Петя курит и ест ягоды
Peter is_smoking and is_eating berries

Datapoint 1: {OBJ} {is smoking; berries (OBJ)}
Feature 3: 0.06
Datapoint 2: {SUB} {is_eating; Peter (SUB)}
Feature 3: 0.80
Training the guesser: probabilistic features

Feature 4: the probability of an argument in a given position being adopted

<table>
<thead>
<tr>
<th>Position of a potential adoptee</th>
<th>Adoption rate (all arguments)</th>
<th>Adoption rate (SUBs only)</th>
<th>Adoption rate (non-SUBs only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL (on the first conjunct and to the left of it)</td>
<td>0.75</td>
<td>0.93</td>
<td>0.17</td>
</tr>
<tr>
<td>FR (on the first conjunct and to the right of it)</td>
<td>0.06</td>
<td>0.65</td>
<td>0.02</td>
</tr>
<tr>
<td>M (on a middle conjunct)</td>
<td>0.03</td>
<td>0.21</td>
<td>0.01</td>
</tr>
<tr>
<td>LL (on the last conjunct and to the left of it)</td>
<td>0.01</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>LR (on the last conjunct and to the right of it)</td>
<td>0.03</td>
<td>0.25</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Training the guesser: probabilistic features

Feature 4: the probability of an argument in a given position being adopted

Петя курит и ест ягоды
Peter is_smoking and is_eating berries

Datapoint 1: \{LR\} \{is_smoking; berries (OBJ)\}
Feature 4: 0.03

Datapoint 2: \{FL\} \{is_eating; Peter (SUB)\}
Feature 4: 0.75
Features and rules

- Features 1 and 2 (argument probability; frame probability) are calculated using the whole Syntagrus.
- Features 3 and 4 (argument adoption rate; position adoption rate) are calculated using the training sample.
- In addition, there are several deterministic rules which prohibit certain adoptions (overriding probabilistic features), e.g.:
  - If a potential adopter has person, number, gender or mood different from the potential subject adoptee’s real parent, then the adoption is impossible.
Guessing

The guesser finds an average of the four probabilities. All datapoints with an average higher than a certain cutpoint are considered to be cases of real adoption, others are not. The algorithm finds the optimal cutpoint for the training set by trial and error, and then applies the calculated feature values to the test set.
Guessing

The guesser finds an average of the four probabilities. All datapoints with an average higher than a certain cutpoint are considered to be cases of real adoption, others are not. The algorithm finds the optimal cutpoint for the training set by trial and error, and then applies the calculated feature values to the test set.

<table>
<thead>
<tr>
<th>Rules</th>
<th>Outcome</th>
<th>Ave.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Rules</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>{is_eating; Peter}</td>
<td>yes</td>
<td>0.51</td>
<td>0.30</td>
<td>0.18</td>
<td>0.80</td>
<td>0.75</td>
<td>_</td>
<td>yes</td>
</tr>
<tr>
<td>{is_smoking; berries}</td>
<td>no</td>
<td>0.06</td>
<td>0.13</td>
<td>0.00</td>
<td>0.06</td>
<td>0.03</td>
<td>_</td>
<td>no</td>
</tr>
</tbody>
</table>
Guessing

The guesser finds an average of the four probabilities. All datapoints with an average higher than a certain cutpoint are considered to be cases of real adoption, others are not. The algorithm finds the optimal cutpoint for the training set by trial and error, and then applies the calculated feature values to the test set.

<table>
<thead>
<tr>
<th>{is_eating; Peter}</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Ave.</th>
<th>Rules</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.30</td>
<td>0.18</td>
<td>0.80</td>
<td>0.75</td>
<td><strong>0.51</strong></td>
<td>_</td>
<td>yes</td>
</tr>
<tr>
<td>{is_smoking; berries}</td>
<td>0.13</td>
<td>0.00</td>
<td>0.06</td>
<td>0.03</td>
<td><strong>0.06</strong></td>
<td>_</td>
<td>no</td>
</tr>
</tbody>
</table>

From this toy example the guesser will learn:
F3 values; F4 values; cutpoint=0.06
III. WHAT DO WE GET?
**Results** (5-fold cross-section validation with rule-based correction)

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>F-score</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.97</td>
<td>0.92</td>
<td>0.95</td>
<td>0.89</td>
</tr>
<tr>
<td>SUB</td>
<td>0.95</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>OBJ</td>
<td>0.96</td>
<td>0.31</td>
<td>0.40</td>
<td>0.28</td>
</tr>
<tr>
<td>OBL</td>
<td>0.97</td>
<td>0.21</td>
<td>0.20</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Evaluation and error analysis

• False positives are less frequent that false negatives (false positive rate is 0.01, false negative rate is 0.11), which is good
• More than half of the cases identified by the guesser as false positives at an intermediate work stage turned out to be human annotation errors, which means...
• ...that the guesser can be used an error-correction tool. We have already managed to identify some missing shared dependents in the OCS part of the TOROT corpus
Evaluation and error analysis

- The results are excellent for SUBs, but poor for OBJs and OBLs due to very low number of positive examples (resp. 16 and 8)
Evaluation and error analysis

• The results are excellent for SUBs, but poor for OBJs and OBLs due to very low number of positive examples
• Increasing the number of positive examples will boost the performance. Indeed, a preliminary test conducted on a larger sample confirms that:

<table>
<thead>
<tr>
<th></th>
<th>“Current” sample (1103)</th>
<th>“New” sample (2468)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Examples</td>
<td>F-score</td>
</tr>
<tr>
<td>OBJ</td>
<td>16</td>
<td>0.31</td>
</tr>
<tr>
<td>OBL</td>
<td>8</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Conclusion

• We presented a solution for a specific technical issue
Conclusion

• We presented a solution for a specific technical issue
• ...which, however, can be of wider importance for dealing with various coordination-related issues in treebanks
Conclusion

• We presented a solution for a specific technical issue
• ...which, however, can be of wider importance for dealing with various coordination-related issues in treebanks
• ...and for understanding (Russian) coordination, both as an NLP challenge and a linguistic phenomenon
Linguistic insights about coordination

Are *topical* elements more likely to be shared?

<table>
<thead>
<tr>
<th>Position of a potential adoptee</th>
<th>Adoption rate (all arguments)</th>
<th>Adoption rate (SUBs only)</th>
<th>Adoption rate (non-SUBs only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL (on the first conjunct and to the left of it)</td>
<td>0.75</td>
<td>0.93</td>
<td>0.17</td>
</tr>
<tr>
<td>FR (on the first conjunct and to the right of it)</td>
<td>0.06</td>
<td>0.65</td>
<td>0.02</td>
</tr>
<tr>
<td>M (on a middle conjunct)</td>
<td>0.03</td>
<td>0.21</td>
<td>0.01</td>
</tr>
<tr>
<td>LL (on the last conjunct and to the left of it)</td>
<td>0.01</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>LR (on the last conjunct and to the right of it)</td>
<td>0.03</td>
<td>0.25</td>
<td>0.02</td>
</tr>
</tbody>
</table>