

# Methods for Semantic Role Labeling of Russian Texts

A. Shelmanov, I. Smirnov

[www.isa.ru](http://www.isa.ru)  
[www.exactus.ru](http://www.exactus.ru)  
[expert.exactus.ru](http://expert.exactus.ru)



# Task of semantic role labeling (SRL)

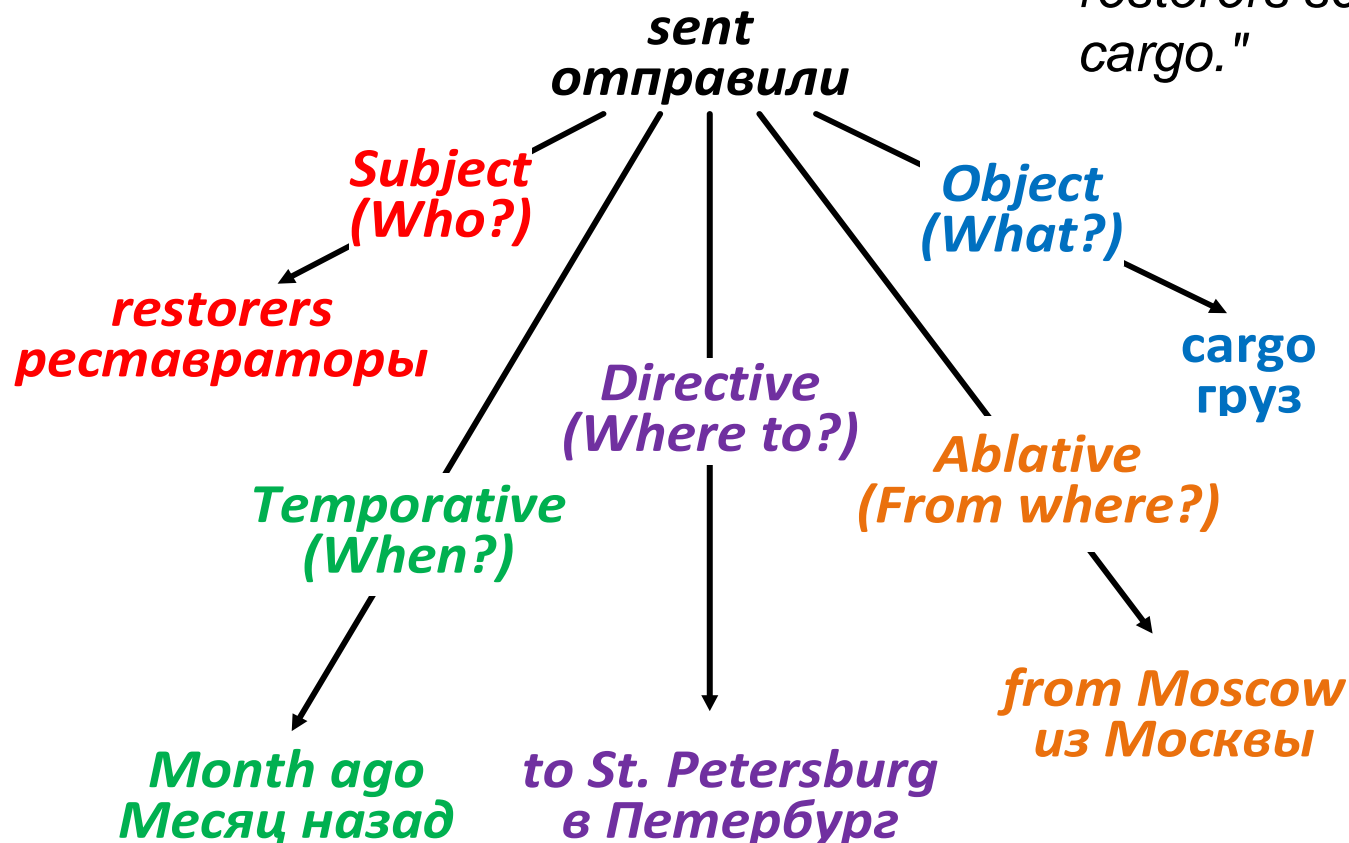
## Semantic role labeling

(shallow semantic parsing) :

1. Determines situations in sentence
2. Identifies arguments of situations
3. Classifies arguments of situations and assigns them thematic roles

Example: "Месяц назад из Москвы в Петербург столичные реставраторы отправили необычный груз"

"Month ago from Moscow to St. Petersburg, metropolitan restorers sent an unusual cargo."



# Practical usage of SRL

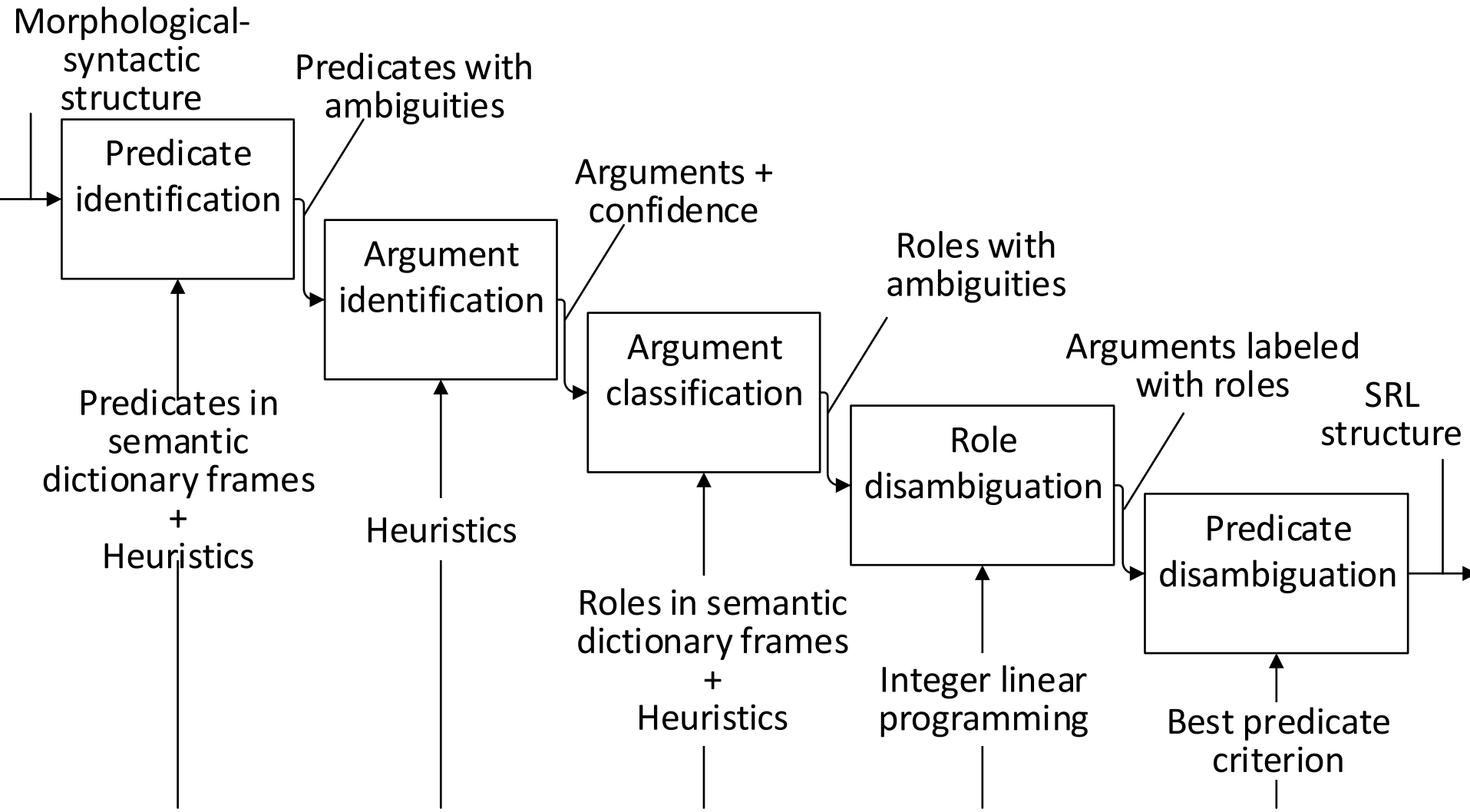
- Question-answering search
- Information extraction
- Information search
- Summarization
- Machine translation

# SRL using semantic dictionary (1)

- Semantic dictionary contains 2,856 frames and 3,585 predicate words (verbs, verbal nouns) for Russian
- Frame for predicate words: “send”, “guide”, ...

<b>Semantic role</b>	<b>Categorial class</b>	<b>Preposition</b>	<b>Grammar Case</b>
<b>Ablative</b>	Location, concrete	Из, из-за, из-под, от, с	Genitive
<b>Addressee</b>	Personal	К	Dative
<b>Destinative</b>	Any	Для	Genitive
<b>Directive</b>	Location, concrete	В, за, на	Accusative
<b>Mediative</b>	Concrete	По	Dative
<b>Object</b>	Any	–	Accusative
<b>Objective</b>	Any	За	Instrumental
		На	Accusative
<b>Subject</b>	Personal	–	Nominative

# SRL using semantic dictionary (2)

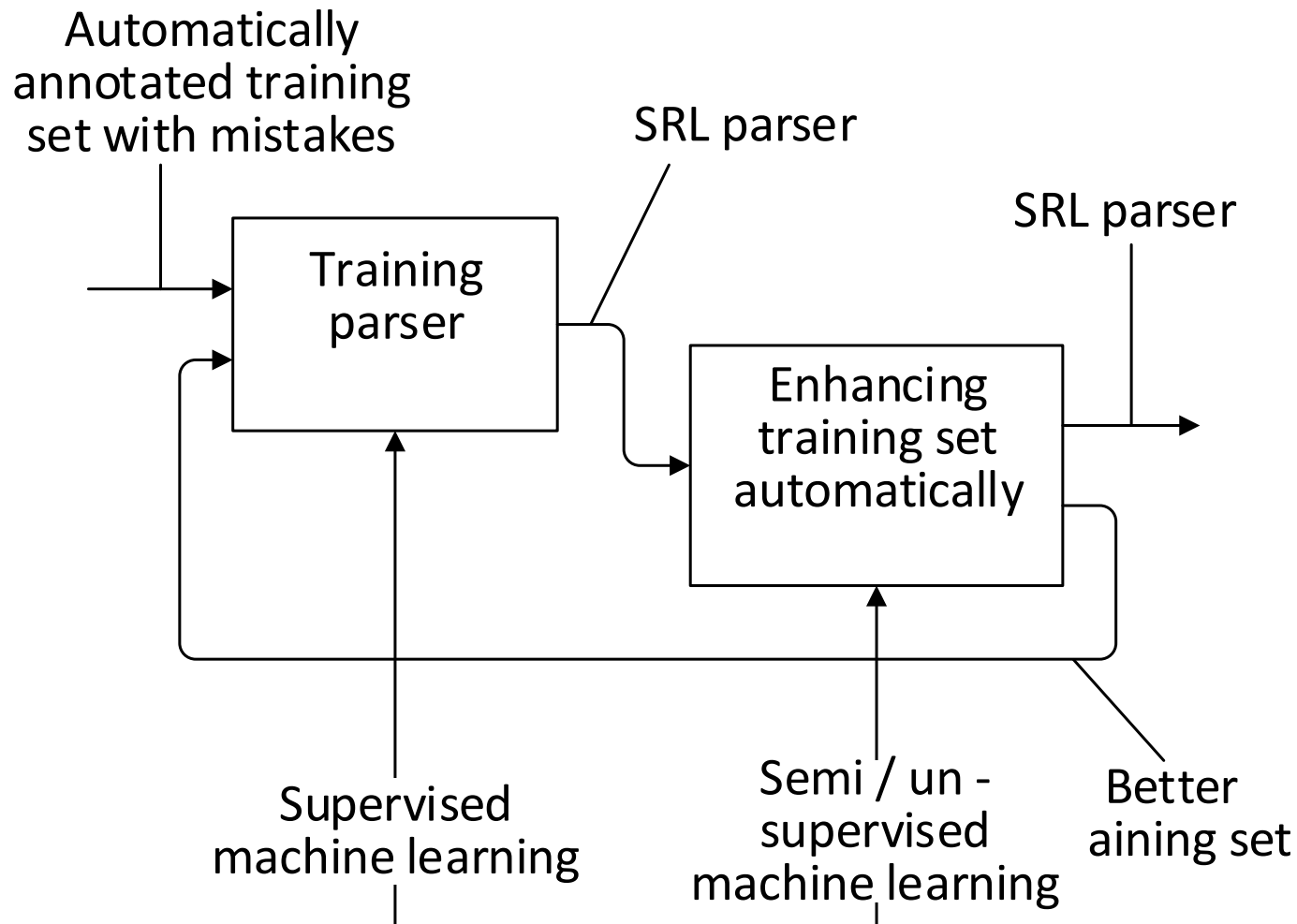


# Why we need machine learning in SRL

- Dictionary-based parser is too sensitive to mistakes in morphological and syntactic annotation
- Heuristics of dictionary-based parser depend on morphological and syntactic annotation schemes
- Need automatic expansion of annotations to new domains

# Training corpus

- Annotated 48,000 sentences of SynTagRus automatically with dictionary-based parser
- Enhancing training set automatically



# Data-driven SRL parsing

1. Joint semantic-syntactic parsing using MaltParser (Nivre, et al)
  - Predicts transitions using linear classifier
  - Labels arcs
  - Complexity:  $O(\text{sentence length})$
2. Data-driven argument classifier
  - Predicts labels of identified arguments using linear classifier
  - Utilizes assessments of classifier for role disambiguation



# Test corpus and evaluation measures

- Test corpus with SRL annotation:
  - Subset of SynTagRus manually annotated with roles: 1,730 sentences (29,041 tokens without punctuation)
  - 3,871 tokens have semantic roles, 61 roles are unique
  - But still contains mistakes and not all examples are annotated completely
- Measured precision, recall, F1, but only on annotated examples

**Dictionary-based parser**

Test set	Recall, %	Precision, %	F <sub>1</sub> , %
All features are GOLD	82.5	94.5	88.1
Gold morphology + dirty syntax and CSC	67.3	88.7	76.5

**SRL using MaltParser**

Test set	Recall, %	Precision, %	F <sub>1</sub> , %
Gold morphology + dirty syntax and CSC	64.9	87.3	74.5

**SRL using MaltParser with training set enhancement**

Test set	Recall, %	Precision, %	F <sub>1</sub> , %
Gold morphology + dirty syntax and CSC	66.3	87.7	75.5

## Dictionary-based parser

Test set	Recall, %	Precision, %	F1, %
All features are DIRTY	65.9	85.7	74.5

## Data-driven parser

Test set	Recall, %	Precision, %	F1, %
All features are DIRTY	66.5	85.4	74.8

## Data-driven parser with training set enhancement

Test set	Recall, %	Precision, %	F1, %
All features are DIRTY	66.7	85.9	75.1

# Methods for Semantic Role Labeling of Russian Texts

A. Shelmanov, I. Smirnov

[www.isa.ru](http://www.isa.ru)  
[www.exactus.ru](http://www.exactus.ru)  
[expert.exactus.ru](http://expert.exactus.ru)

