



Санкт-Петербургский  
государственный  
университет

# Semantic analysis: theory, applications and use cases

6th Seminar of FRUCT

Dmitry Kan, Vladimir Poroshin

# Road map

- Math model of a Natural Language (NL)
- Three levels of text analysis
- Semantics vs syntax
- Applications
- Use cases

# Math model of a Natural Language

- Backbone of NL: verbs + prepositions
- Verb =  $F(\text{arg}_1, \dots, \text{arg}_n)$
- Prepositions: 3D space (*behind*), time space (*during*), cause-and-effect relation (*due to*)
- Class hierarchy (=world picture)
- Basis functions
- Words adjunction

# Math model: basis functions

- $\text{Caus}(x,y) = x \text{ causes } y$
- $\text{Cont}(x) = x \text{ continues}$
- $\text{Hab}(x,y) = x \text{ has } y$
- $\text{Incep}(x) = x \text{ begins}$
- $\text{Oper}(x,y) = x \text{ performs } y$
- $\text{Lab}(x,y) = x \text{ under action of } y$
- $\text{Prepar}(x) = \text{prepare } x, x \text{ is prepared}$
- $\text{Fin}(x) = x \text{ finishes, stops}$

# Math model: examples

- Caus(Subj, Fin Lab(Accus, FIRE)) = to extinguish a fire (=cause to stop having Subj under action of fire)
- Caus(Subj, Prepar(FOOD Accus)) = to stew the vegetables (cause the Subj to get cooked)
- Both examples map to the same Russian verb "тушить" => semantic disambiguation

# Road map

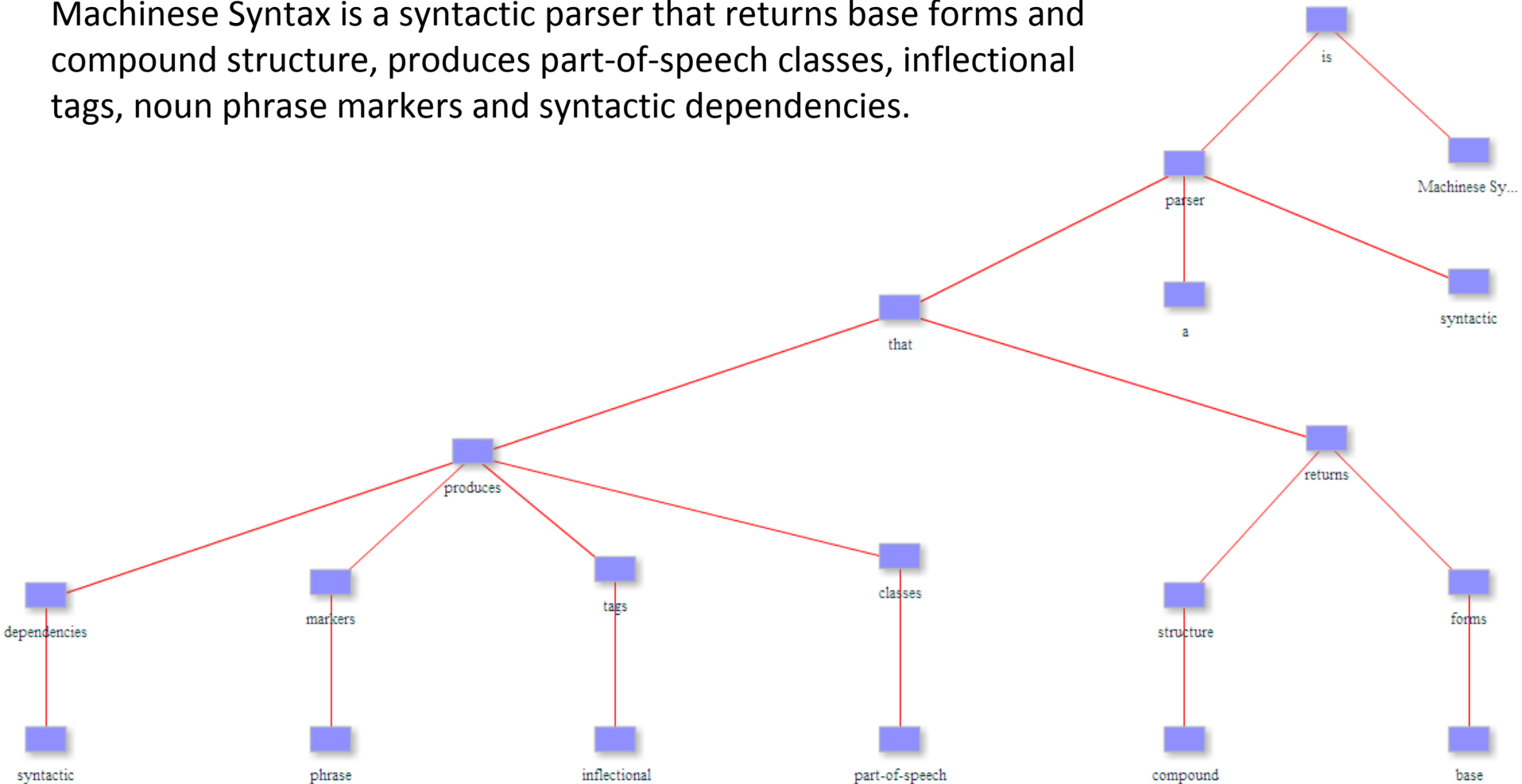
- Math model of a NL
- **Three levels of text analysis**
- Semantics vs syntax
- Applications
- Use cases

# Three levels of analysing text

- Morphological analysis: word level
- Syntactic and semantic analysis: sentence level
- Object properties and relationships, anaphora resolution: text level
- Sentence =  $P(f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n))$  – superposition of functions

# Three levels of analysing text

Machinese Syntax is a syntactic parser that returns base forms and compound structure, produces part-of-speech classes, inflectional tags, noun phrase markers and syntactic dependencies.



# Road map

- Math model of a NL
- Three levels of text analysis
- **Semantics vs syntax**
- Applications
- Use cases

# Semantics vs syntax

- Он пришёл **из** вежливости (He came out of courtesy) **WHY?**
- **Из (вежливость)** generates "WHY?"
- Он пришёл **из** деревни (He came from a village) **WHERE FROM?**
- **Из(деревня)** generates "WHERE FROM?"

# Road map

- Math model of a NL
- Three levels of text analysis
- Semantics vs syntax
- **Applications**
- Use cases

# Applications

- Intellectual search systems
- Question-answering systems
- Spell checker
- Summarization
- Sentiment analysis
- Machine Translation
- Knowledge base
- Facts extraction

# Road map

- Math model of a NL
- Three levels of text analysis
- Semantics vs syntax
- Applications
- **Use cases**

# Use cases

- Smart street dating: semantic search of the best matching candidates around you
- FAQ mobile agent: automatically suggest solutions to the support requests
- Sentiment recognition: goods evaluation
- Automatic summarization: limit information load on mobile devices

# References

- [1] Tuzov V. A.: Computer semantics of Russian language. Saint-Petersburg University Press, Saint-Petersburg, 2004 (in Russian).
- [2] Kan D. A., Lebedev I.S.: Method of formalizing semantical links between objects in a natural language text. Bulletin of Saint-Petersburg University. Series 10. 2008. Issue 2. pp 56-61 (in Russian).
- [3] Kan D. A.: Method for automatic creation of translational semantic dictionary for Machine Translation // XL Conference Control Processes and Stability'09, pp. 429-435 (in Russian).
- [4] [www.semanticalyzer.info](http://www.semanticalyzer.info)

# Questions



Thank you!

