

Linguistic Semantics for Search Precision and Recall Improvement (part 1)

Ilya Tikhomirov

Introduction

The first lecture gives a general introduction to Linguistic Semantics, its brief history and definitions of natural language processing levels (lemmatizing, morphological, syntactic and semantic analysis). The introduction to Communicative Grammar is given.

What does linguistic semantics mean?

Linguistic semantics is language independent and operate with invariant meanings.

Semantic analysis is the process of relating syntactic structures from the levels of phrases, clauses, sentences to the level of language-independent meanings.

History (1)

Cambridge linguistic society: language-mediator (1959).

The task of language-mediator is language-independent semantic description. Language-mediator contains classifier of meaning names and rules for formula building.

Disadvantages: invisibility of situation, described by patterns. Lexical and semantic penury, absence of automatic translation of the natural language text to the language-mediator.

History (2)

Noam Chomsky: transformational grammar (1962).

Mechanism for natural language sentence generation. Extension of the theory lead to taking into account of the additional semantic information about the words. It's allow to solve the next tasks:

- define meanings of generated sentences;
- semantic anomaly finding;
- establishment of the sentences equivalence.

Disadvantage – transformational grammar is generative.

History (3)

Richard Montague: Formal semantic (1974).

Montague thesis - natural languages and formal languages (like the language of logic) can be treated in the same way.

The Montague grammar is based on:

- higher order predicate logic;
- lambda calculus;
- notions of intensional logic.

The main problem: difficulty of natural language automatic translation into predicate logic.

History (4)

Charles Fillmore: Case Grammar (1968).

Case Grammar focuses on the link between the valence of a verb and the grammatical context it requires. This theory analyzes the surface syntactic structure of sentences by studying the combination of deep cases (semantic roles):

- agent (A);
- object (O);
- benefactor (B);
- instrument (I).

Each role is required by a specific verb. Example:

- Jones (A) gave money (O) to the school (B).

The influence of Case grammar on linguistics has been significant.

History (5)

I.A. Melchuk: Meaning-text model (1974).

Main task of the model is description of natural language and transition from text to the meaning and back. Features of the model:

- hierarchic structure (different levels of representation : text level, morphology-shallow, morphology-deep, syntactical-shallow, syntactical-deep, semantic);
- each level has own elements, representation rules and transition rules to others levels;
- original syntactical and lexical functions theory;
- original explanatory-combinatory dictionary.

The model has theoretical character, but have some applications (for example, ETAP-3 translator).

History (6)

G.A. Zolotova: Communicative grammar (1982).

Communicative grammar is an original approach to syntactical and semantic analysis.

- It disproves opposition of syntax and semantic.
- Communicative Grammar is explanatory grammar (in contrast to descriptive grammars).
- Analysis rules should take into account categorical words meaning.
- Main element of the text is syntaxeme (minimal indivisible semantic-syntactic structure of language).

From word meaning to text meaning

WORD -> SENTENCE -> TEXT -> LIFE ☺
meaning meaning meaning meaning

Text Analysis (1)

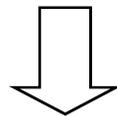
Levels of machine text analysis:

- Lemmatizing.
- Morphological analysis.
- Syntactic analysis.
- Semantic analysis.

Lemmatizing

The main task of the Lemmatizing analysis is to recognize words and separators in the text.

The mother brings her son to school.

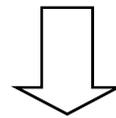


the	mother	brings	her	son	to	school
-----	--------	--------	-----	-----	----	--------

Morphological analysis

The main task of the Morphological analysis is to define part of speech (verb, noun...) and form of words (case, singular/plural form, gender for nouns ...).

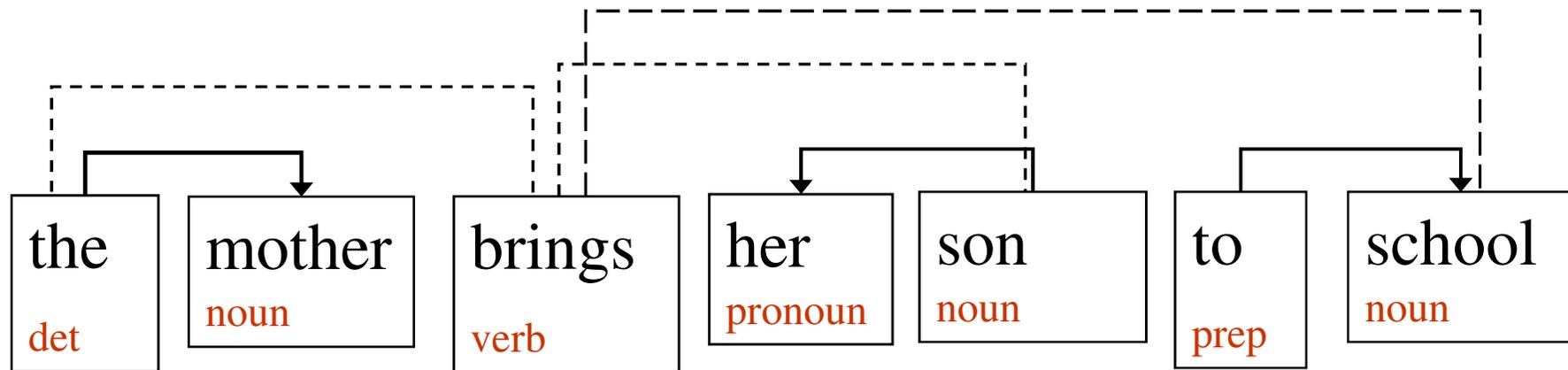
the mother brings her son to school



the	mother	brings	her	son	to	school
det	noun	verb	pronoun	noun	prep	noun

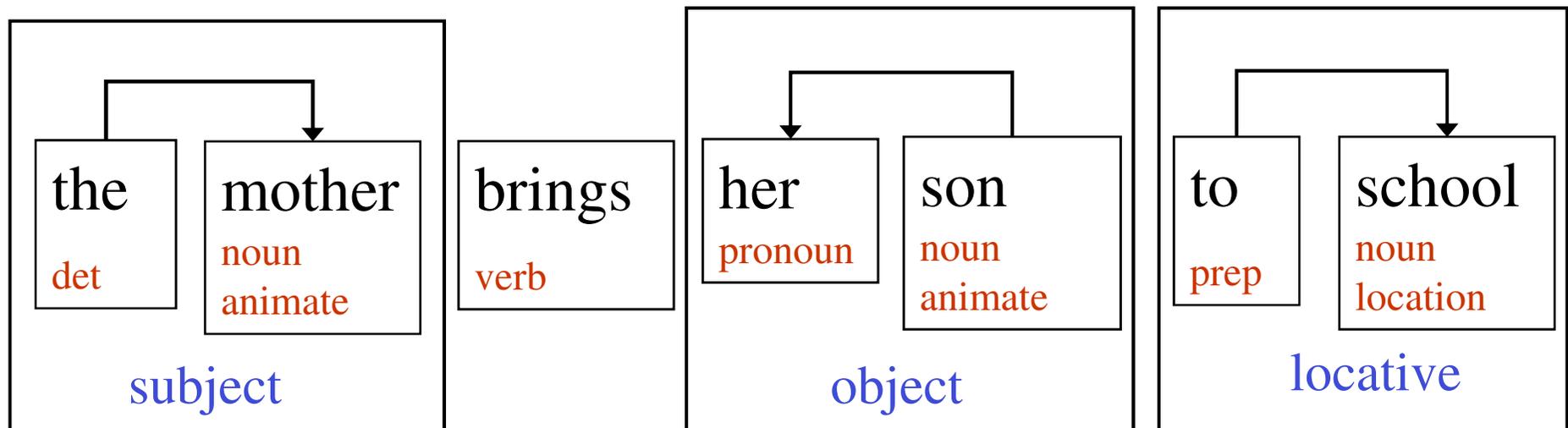
Syntactic analysis

The main task of the syntactic analysis is to establish syntactic dependencies between lexemes defined at the previous stage.



Semantic analysis

The main task of the semantic analysis is to reveal semantic meanings.



Communicative Grammar

One of the main terms in Communicative grammar is **syntaxeme**.

Syntaxemes are detected taking into account:

- categorial semantics of the word;
- morphological form;
- function in the sentence.

Communicative Grammar: categorical semantics (1)

Let's start from example:

- The lecturer delivers a speech.
- 1) It's easy to replace «the lecturer» by «the professor» or by «the student». The sentence keeps the meaning.
 - 2) But if we try to replace «the lecturer» by «the sofa» or by «the time» or by «the dozen», the sentence becomes meaningless. **WHY?**

In the first case we mean animate person and replace person by person. In the second case we try to replace person by inanimate word or quantitative word. So, we use words from others categorical semantic classes.

Communicative Grammar: categorical semantics (2)

Categorical semantic class is
generalized meaning of the word,
in contrast to its lexical meaning.

Generalized meaning of the word
defines its syntactical and
semantic abilities.

Communicative Grammar: categorical semantics (3)

Syntaxemes can be similar in morphological form, but different in their meaning:

- 1) The mother brings her son **to school**.
- 2) Laziness leads researcher **to trouble**.

Categorical semantic class is non-morphological category.

Communicative Grammar: function (1)

Types of syntaxeme function (usage in speech):

- 1) independent usage – syntaxeme meaning does not depend on the context (Example: Book title, News title);
- 2) component of the sentence:
 - grammatical subject (**John** is yellow);
 - grammatical predicate (John **is yellow**);
 - grammatical modifier (John is yellow **for the long time**).
- 3) component of the phrase(Phrase is a group of words with a main syntaxeme):
 - verb;
 - adjective;
 - adverb;
 - ...

Communicative Grammar: function (2)

Syntaxemes are divided into three types
(according to their functions in the sentence):

- **free** (can be used in any place of the text);
- **caused** (can be used as a component of the sentence);
- **bounded** (can be used as a component of the phrase).

Communicative Grammar: Simple sentence (1)

Simple sentence is a minimal communicative unit. Any simple sentence contains at least subject and predicate.

Simple sentences can be divided into:

- Monopredicative (Example: John is yellow).
- Polypredicative (Example: Laziness leads researcher to trouble).

Communicative Grammar: Simple sentence (2)

Five models of simple sentence are distinguished:

1. Subject and it's action (John writes the letter).
2. Subject and it's state (I am cold).
3. Subject and it's property (John is brave).
4. Subject and it's quantity (Two brothers).
5. Subject and it's classifier (John is a student).

Any complex sentence can be represented as a set of simple sentences.

Communicative Grammar: meaning of the sentence

Meaning of the sentence is defined by a set of syntaxeme's meanings and relations on a set of syntaxemes.

Meanings examples:

- Subject – main actor in the sentence;
- Temporative – component denotes time or date;
- ... (near 85).

Communicative Grammar: conclusion

Communicative grammar features:

- explanation grammar;
- comprehended discourse focus;
- meaning oriented;
- formalized structure.

Communicative grammar is applicable for the search tasks.

Contacts

Institute for Systems Analysis
Russian Academy of Science
117312, Moscow,
pr. 60-let Octiabrya, 9
phone/fax: +7 (499) 135-04-63
e-mail: matandra@isa.ru