COGNITIVE MODEL OF TIME AND ANALYSIS OF NATURAL LANGUAGE TEXTS

Xenia A. Naidenova, Marina I. Garina.

***Abstract**: The extension to new languages is a well known bottleneck for any text analyzing system. In this paper, a cognitive model of time is proposed and the questions of extracting events and their time characteristics from texts are discussed. The cognitive model of time due to its independence of concrete natural language can be considered as a basis for constructing text mining systems intended for extracting temporary relations.*

***Keywords**: Natural Language Processing, cognitive model, time model.*

***ACM Classification Keywords**: I.2.7. Computing Methodologies - Artificial intelligence - Natural Language Processing.*

Introduction

Time representation and reasoning is an issue of many different disciplines. Studies in this area exploit several sources: cognition, language, perception as well as world knowledge and the difficulties of these studies are explained by the inherent complexity and multidimensionality of time [Elkin, 2008] as a human thinking category. Modern psycho-linguistic and neuro-linguistic investigations show that mechanisms of thinking and mechanisms of thinking verbalization are different from one another [Popova, & Sternin,2007]. But human minds have the ability to establish systematic relationships between linguistic forms and perceptually based knowledge. Kubrjakova E.S. [2004] has formulated a new cognitive – discursive investigation direction, insisting on the thesis that each language phenomenon can be adequately described and explained only if it has been studied in the framework of cognition and communication. A goal of cognitive linguistic is to find adequate cognitive construction for every language form. Our perspective aim is creating a cognitive model of time. This model being translated in different natural languages could serve as a basis for text mining and extracting information of temporary attributes of events. Presumably, the text processing system consists of:

- **Cognitive models of time** and events that oriented to a given domain application and the goals of text processing;
- **Translator** that is adjusted to a given NL;
- **Block of plausible (commonsense) reasoning** to infer consequences from established temporary relations between events in the text by means of meta-knowledge of cognitive models;
- **Dialogue Syntactical Analyzer** for a given NL;

- **Block of control** or operational subsystem of the translator.

According to Figure 1, Text interacts with Translator and Syntactical Analyzer; Translator interacts with Syntactical Analyzer and Cognitive Models of Time and Event. As a result, events and their time moments or intervals are extracted from Text, and then the conclusions about temporary relations between events are done.

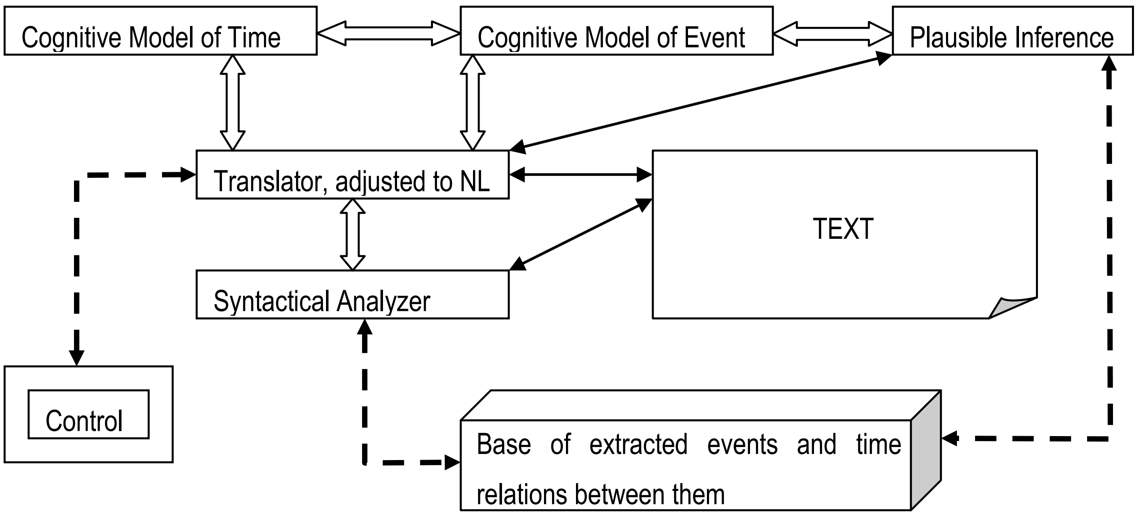


Figure – 1. Approximate structure of the text processing system.

State-of-the-art

The volume of the literary sources related to text mining temporary information about events is enormous. All sources can be divided into two groups: 1) the works developing the logical theory of time or logic of time [Moszkowsky, 2007]; and 2) the works, connected with extracting information about events and their time characteristics from the texts in different natural languages [Boguraev et al., 2006]. Both these directions have some limitations. The former is occupied by the problems of inferring consequences from the facts, already extracted from tests. The latter does not yet have a general platform for representing knowledge about the world and lingual structures, at least in limits of one of the lingual groups. Cognitive simulation is a "bridge", a connecting link, which is necessary for understanding the principles of interaction between knowledge, reasoning, and lingual abilities.

Model of time

The main purpose of time model and applying it to some text is to reveal **events** appearing in this text. Event as a concept can have the name and some other properties (may be empty) and, as a rule, it is associated with **time interval**.

In general, a time interval consists of two markers: (the beginning; the end) and has the duration. But sometimes we need to use intervals opened in the past (without beginning) or in the future (without end) or both.

Time interval can be expressed via some events, for example, "at dawn, to the first volleys of artillery", "long before the first sun rays". Events are used as time markers in this case. Also particular cases of time interval are: unit of time, a set of units of time, a moment of time. It is worth mentioning that the very moment of time can be an event, for example, "September began", "Days go".

The cognitive model of time includes the following elements:

- the **units** of time (year, month, spring, minute);
- the **time intervals** and their properties: the beginning, the end, duration, without beginning (opened in the past), without end (opened in the future), consisting of points;
- **environment**: nearest past, nearest future (about noon, soon after the beginning/the end, toward the evening);
- **various relations** between units and intervals: coincidence, contact, precedence, going after, intersection, inclusion, remoteness into the past/future time;
- **degree of relations**: the measures of remoteness, intersection and so on;
- **comparison relations of interval duration**: longer, less for long, shorter and so on;
- **uncertain (fuzzy) relations**: considerably later, once, early in the morning and so on.

Many references to dates or times in a text are not fully specified, with the result that some parts of extracted knowledge about time and events will have to be computed from the context during the interpretation stage.

In Figure 2, the basic cognitive model of time and the relation between its elements considered above is represented. Obviously, this model is recursive, i.e., an event is associated with a certain time interval, and a time interval can be expressed via some events (look above).

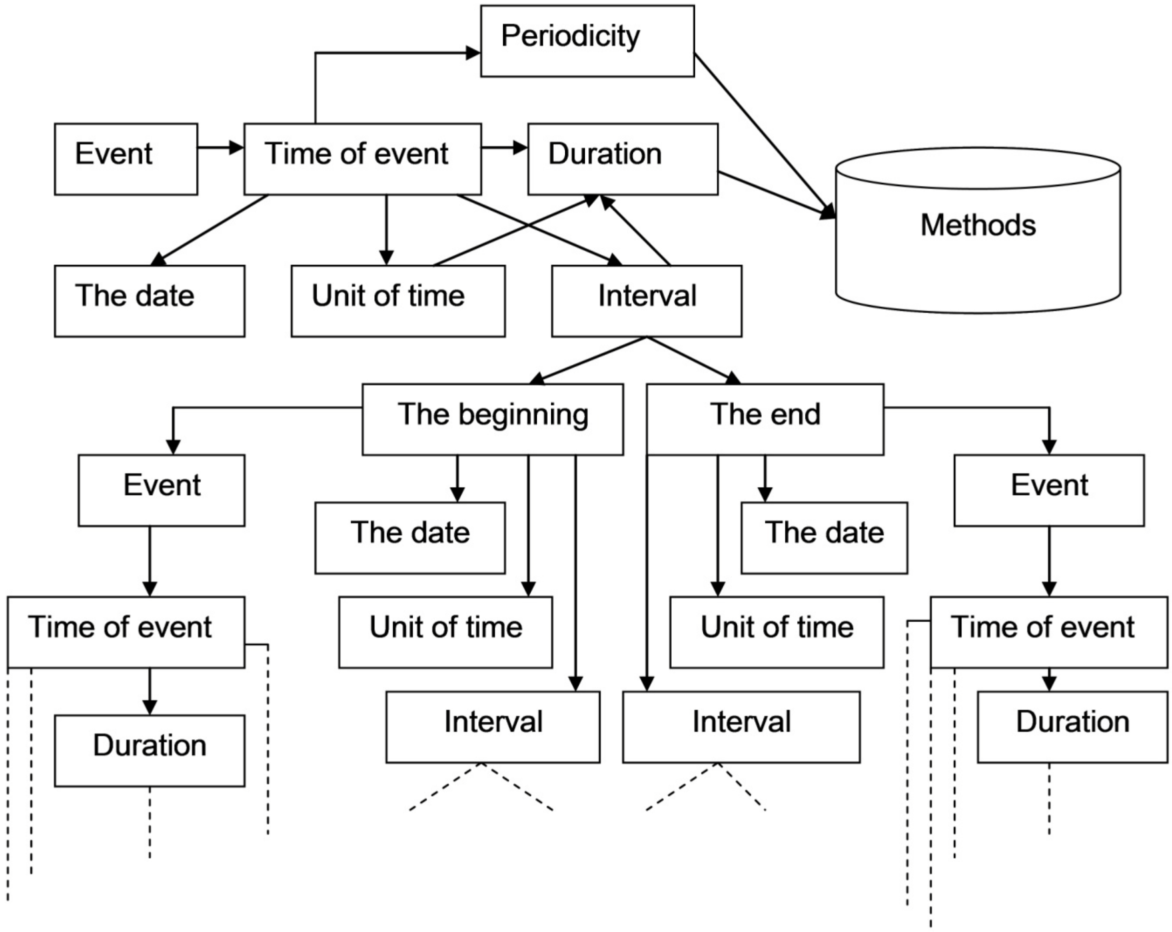


Figure - 2. The Basic Cognitive Model of Time

Methods

At first, it is important to be able to calculate the duration of time interval associated with some event. This duration can be calculated with the following **computation rules**:

- as the difference between the end and the beginning of time interval associated with the precise time markers (the dates);
- via the events being the markers of the beginning and the end of a time interval: <EVENT1> to <EVENT2>;
- as a set of time units (for example, 900 days).
-

Further, it is necessary to demonstrate the truthfulness of different enumerated above **relations between time intervals.** The most useful relations are listed below.

- **Sequence** (which event is earlier, which is later, which will be the next, which already happens). At best we concern a strict or weak order relation. Events in this relation can **contact** (in this case we have null duration between events).
- **Simultaneity** with it's particular cases:
  - the same time interval of events,
  - inclusion one interval into another,
  - intersection of intervals.

So the methods we discuss are natural rules for analyzing the relationships between time units/intervals. They include both **computations** and implicative assertions of the general kind:

$$E1.T.end < E2.T.begin \leftrightarrow E1\ R_{precedence}\ E2.$$

Here E1 and E2 are events, E.T is the interval associated with event E, T.end and T.begin are the end and the beginning of interval T, respectively, and $R_{precedence}$ is the precedence relation. In object-oriented design terms here we deal with access to fields of objects of classes Event and Time_Interval. Thus it's important to watch for the following conditions' performance:

- T.begin ≤ T.end or T.end ≥ T.begin with the precise time markers;
- T.begin $R_{precedence}$ T.end if events are used as markers.

Next, it may be useful to detect some properties of events appearing in the text, such as:

- Fuzziness. Events can be indeterminate (fuzzy) in the time (considerably later, once upon a time).
- Speed. Events can flow in the time rapidly or slowly.
- Frequency. Periodic events can be frequent or rare.
- The temporary properties of events can be estimated in both objective and subjective manner.

A set of events can be associated with only one time interval, so those events can be indiscernible that will lead to a weak order relation. An event can be expressed both by only one word and by a set of proposals (maybe only one proposal). So there should be a method to parse complex expressions, especially in view of an event and the time interval associated with it can be in different proposals. It is possible that an active agent (including temporary moment) cannot be determined without the aid of referential relation. It is also necessary to take advantage that there are the events attached by default to the time intervals, such as dawn, sunset, school-leaving ball, dinner, supper, breakfast, the beginning of workday and so on.

Constructing a translator

The cognitive model of time does not depend on language, but is tuned into different natural languages. A translator of the cognitive model of time into language expressions for a given natural language and vice versa can be built as a trained system that learns by specially constructed phrases. For this goal, the following levels of natural language are considered: lexical, morphological, and syntactic ones.

Let's discuss the **lexical level** now. A special type of time interval is the name of time unit, for example: *TI* = {century, year, month, twenty-four hours, the morning, day, evening, night, January, February, March, April, May, June, minute, second, winter, summer…}. There are some banal relations between these units, such as:

- Classification ("is-a"), for example: "seasons are winter, spring, summer, and autumn";
- composition ("consist-of"), for example: "twenty-four hours consist of night and day";
- part-whole, for example: "minute is a part of hour";
- occurring in cycles, for example: "winter of one year follows after autumn of previous year";
- inclusion;
- sequence, for example "spring comes after winter".

In Figure 3, the relation of classification is shown with the aid of triangle connections while the relation of composition is shown by simple arrow. If the relation of composition is determined between the intervals of upper level, then it is determined between the interval-descendants, for example, June consists of twenty-four hours. Specific dimensionality can be determined only for the connections of the lower level. It cannot be said how many twenty-four hours year generally consists of, month generally consists of, but it can be said, how many twenty-four hours the leap year consists of, current year consists of, January consists of, etc.

A generalized model of event can be defined as follows: EVENT = $\langle E, R_e, Pat_e \rangle$, where E is a set of classes of events, R_e is the relationship: $R_e \subseteq E \times E$, $R_e = R_{class} \cup R_{comp}$, i.e., events are also organized into the hierarchy of classification "is- a" (R_{class}) and composition "consist-of" (R_{comp}). Pat_e is a set of the regular expressions (patterns, templates), which make it possible to take out the text candidates into the exemplars of events of each class. Each candidate can have some parameters (contextual properties), according to which the relevancy of candidate can be evaluated. After establishing a certain threshold value, it is possible to select only the candidates with high probability of being some events.

The search for the beginning and the end of an event can also be achieved with the aid of templates.

Knowing time intervals, associated with events, it is possible, being guided by the ordering relations, determined for the intervals, to establish the same relations for events too.

At lexical level, it is possible to take advantage of the special words, helping to reveal events and their environment. For example, in [Elkin et al., 2008], the groups of words reflecting so-called event-related time are given:

- multiple repetition of one and the same event can be revealed with such words as “daily, every week, quarterly, monthly, and yearly”;
- single event can be revealed with the words “once only, one time, once”;
- such word as “momentary, prolonged” can help to distinguish long and short duration of an event;
- the time before an event is accompanied by the following words: before, in advance, in good time, previously, before the appointed time, it is preliminary, it is premature, on the threshold of, it is earlier than, long before, thus far not, not in a long time, recently, as long as, the day before;
- the time after the event is accompanied by the following words: later, afterward, it is later, then, after, hence, hence-forward, forth, in future, from now on, after all, immediately afterward, further, when;
- these words establish event-related time or time attached to a concrete event: in one’s life, from birth, originally, while;
- affirmation time is revealed with the words: sometimes, someday, in the course of time, then, once;
- “negative” time is revealed with the words : never.

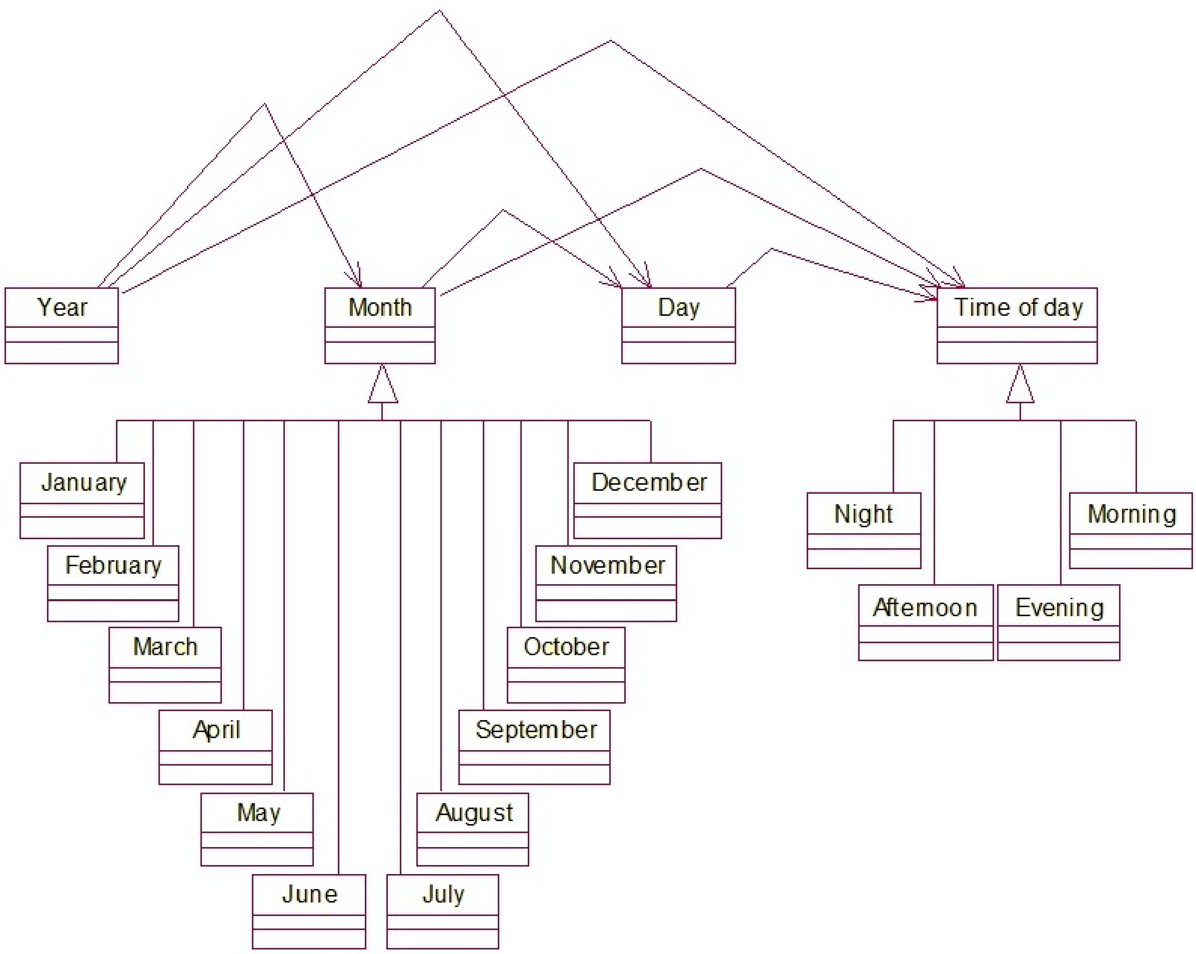


Figure - 3. Fragment of classification relation between time units

In [Kreydlin, 1997], the approximate classification of Russian temporary pretexts has been given (see, please, Table 1).

Table - 1. An example of temporary pretexts.

Relation	Pretext	Example	Temporary marker, the event
Simultaneity Extent Duration	For, During	For entire trip he said nothing During this year, he lives in town	The time interval is attached to event “trip” This year, the time interval is determined event “he lives”
Precedence	Approximately	We awaited approximately to midnight	Temporary marker: midnight; Time: the indeterminate half-interval. Event: we awaited

As to **syntactic level**, it is necessary to develop and use a set of syntactic patterns, such as Table 2 shows. Some principles of extracting syntactic patterns from texts are discussed in [Cimiano, 2006].

Table - 2. Some examples of syntactic patterns.

Relation	Structure	Syntactic pattern	Role in the sentence
Going after	Immediately afterward <action>/<event>, <event>	Adverb with the pretext «afterward» <action>/<event>, <event>	Adverbial modifier of time. Example: Immediately after the wedding and the parting words of parents, they left.
Inclusion	Including <the date>	Verbal Adverb <the date>	Adverbial modifier of time. Example: Including 2010

According to all above-stated translator work consists of following stages:

- At first translator searches for the supporting key words (time markers), which are associated with the expression of time in the text.
- Then translator, using lexical and syntactic models, attempts to determine the events, associated with the chosen time markers.
- If it is necessary, then the Syntactic Analyzer (Parser) is started.

Translator can repeatedly be turned first to the text or first to the cognitive model, then to the syntactic analyzer in order to search purposefully for the required (according to the rules of cognitive model) linguistic constructions.

While translator works assumptions about the events, extracted from the text, with their time intervals become. They can be represented as the list of the possible facts. Then the **Base of Events** will be filled up with the copies of events with their time characteristics. The **Block of Plausible Reasoning** also derives all consequences of the discovered facts (events, their properties, the relations between them).

Example

An example of text analysis is given on the narrative of V. Nekrasov "In the trenches of Stalingrad". This example shows the result of the event-temporary text analysis with the use of cognitive models of time and events. There are some numbered proposals below, chosen from the source text. Some proposals without clearly time markers are passed.

1. For all my life, I can not recollect similar autumn. 2. September passed. 3. In the mornings, fish laps in Volga, and the big circles disperse on the mirror surface of the river. 4. At dawn, to the first volleys of the artillery ... it [the left shore of Volga river] is gentle …... 5. Some time it [the fog] still keeps over the river ……. 6. And long before the first sun rays, the first long-range gun shoots. 7. So the day begins. 8. Exactly at seven o'clock, at first sight imperceptible, the "frame" appears high in the sky. 9. It [The first ten of aircrafts] will determine the entire day.

The Table 3 illustrates the result of extracting time moments and associated events appearing in the text. At first the keywords definitely connected with the indication of time are revealed. The Syntactic Analyzer, Translator, and Cognitive Models are used for obtaining complete information about the events associated with the time indicators, calculating the duration of time intervals, and inferring all the consequences from the facts discovered.

Table - 3. The result of extracting events appearing in the text.

№	Event	Time interval	Inferred information
1	I do not recollect	Autumn	Autumn consists of «September, October, and November».
1	For all my life	The life of the author	
2	September passed	September	It precedes "October"; consequently, "October began"
3	Event 1: Fish laps in Volga; Event 2: The circles disperse	In the mornings	Each day in the morning; October;

	on the surface of the river		
4	X is gentle; X = the left shore of Volga river	At dawn; To the first volleys of the artillery;	At dawn = early in the morning; To (before) the first volleys of the artillery; **Event** = the first volleys of the artillery;
4	The first volleys of the artillery	At dawn;	Early in the morning; October; Autumn.
5	Event: X keeps; X = the fog	Some time	Some time, For a while; Early in the morning.
6	The long-range gun shoots	Long before the first sun rays	Before dawn
6	Event: First sun rays	At dawn	Early in the morning; October; Autumn
7	The day begins	The day	Day comes after morning; The beginning of the day;
8	The «frame» appears	At seven o'clock	At seven A.M.; the beginning of the day.
9	It will determine	The entire day	
9	It = the first ten of aircrafts	The entire day	The entire day = from the morning to the evening

In the second proposal, "September" is the subject. Predicate is expressed by the verb of passed time, whose semantics speaks that the time interval is finished, it left into the past. It is derived from the cognitive model of time that October goes after September, next month of autumn.

The subject in the fourth proposal is established with the aid of referential analysis of the previous proposal.

There is no explicit indication of time moment in the sixth proposal, but adverbial modifier of time «before the first sun rays» is associated with the dawn and the dawn – with the morning. That's why we extract the event «the first sun rays» and associate it with «early in the morning». The proposal «It will determine the entire day» requires returning to the previous proposal in order to associate the word «it» with «the first ten of aircrafts». This action requires the complete syntactic analysis of previous proposal.

The analysis of the text results in obtaining the following sequence of the events: 1) the long-range gun shoots before dawn; 2) the first volleys of the artillery at dawn; 3) the day begins; 4) The «frame» appears at seven A.M.

Conclusion

The proposed model and methods are quite suitable for extracting events and their time characteristics from the text. For achievement of this purpose, the dialogue between the cognitive model of time, the translator and the syntactic analyzer is indispensable. It is necessary to note that the completeness and accuracy of the extracted knowledge depend on the cognitive model of time, its completeness and accuracy.

Subsequently it's necessary to work out the Cognitive Model of Time as completely as possible including all its elements, relations and methods. It is also necessary to take into account the uncertainty of time intervals. The Cognitive Model of Event depends greatly on the field of application. But this model contains also some universal cognitive elements: fact, process, action, result, subject, object, place of event, time of event, causal links between events and properties of object (subject). It's planned by us to refine the Cognitive Event Model through the knowledge of a concrete domain application (business, finances). It implies incorporating a mechanism of plausible inference over events into this model.

The next step of our project is to create the translator as a system of trained links between cognitive structures of time and events and correspondent patterns reflecting these structures in the natural language texts. The cognitive models are based on the knowledge about the world and therefore they can perform a semantic control of the Syntactic Analyzer's activity. However the translator's construction is the object of our further work.

Acknowledgements

The authors thank Doktor Vladimir Polutin, Director of HP Labs Russia, for the initiation and support of this work

Bibliography

[Boguraev, & Ando, 2005] B. Boguraev and R.K. Ando. Time ML-Compliant Text Analysis for Temporal Reasoning. Proceedings of International Joint Conference on Artificial Intelligence (IJCAI'2005), pp. 997-1003. 2005.

[Boguraev, & Ando, 2006] B. Boguraev, R. Minoz, and J. Pustejovsky (Eds). ARTE: Annotating and Reasoning about Time and Event. Proceedings of the Workshop, Sydney, Australia: the Association for Computational Linguistics (ACL), 2006.

[Cimiano, 2006] P. Cimiano. Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. Springer Science+Business Media, New York, 2006, 347 pages.

[Elkin, et al., 2008] S.V. Elkin, V.V. Kulikov, E.C. Klishinsky, O.Y. Mansurova, V.Y. Maksimov, T.N.Musaeva, and S.N. Alineva, The Multi-model Character of Time and Temporal Relations in Semantic Language (SL). Science-Technical Information, Number 2, pp. 18-30, 2008.

[Kreydlin, 1997] G.E Kreydlin. Time through the Prism of Temporary Pretexts // Logical Analysis of Language. Language and Time, Moscow, 1997, 352 pages.

[Kubrjakova, 2004] E. S Kubrjakova. About the Purposes of Cognitive Science and Actual Problems of Cognitive Linguistic. The questions of cognitive linguistic – 2004, number 1, pp.6-17.

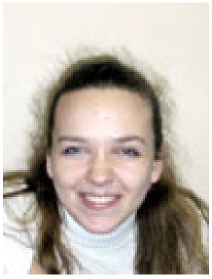
[Moszkowsky, 2007] Ben Moszkowski. Using Temporal Logic to Analyze Temporal Logic: A Hierarchical Approach Based on Intervals. Journal of Logic and Computation, 17(2):333–409, 2007. STRL Publication 2007-8.

[Popova, & Sternin,2007] Z.D. Popova and I.A. Sternin. Cognitive Linguistics. – M.: AST: «East – West». 2007.

Authors' Information

Xenia A. Naidenova – *Military medical academy, Saint-Petersburg, Stoikosty street, 26-1-248, e-mail: ksennaidd@gmail.com*

Major Fields of Scientific Research: Natural Language Processing, Machine Learning, Common-sense Reasoning



Marina I. Garina – *State university of transport ways, 190031, Saint-Petersburg, Moscow Avenue, 9;*

e-mail: MIGarina@gmail.com

Major Fields of Scientific Research: Multi-criteria Decision Aid, Natural Language Processing, Graph Theory