Automatic Processing of Temporal Expressions in Serbian Natural Language Texts

Jelena Jaćimović

University of Belgrade, Faculty of Philology
Studentski trg 3, Belgrade, Serbia
{jjacimovic@rcub.bg.ac.rs}

Abstract. This paper provides an overview of the most important existing resources and computational approaches used for recognition and normalization of temporal expressions. The survey of previous contributions in the field is followed by a description of the ongoing development of a system for recognition and normalization of temporal expressions in Serbian texts, which is the most complete approach that is currently available for Serbian language.
One of the main contributions of this rule-based system is the ability to automatically identify and annotate different types of temporal expressions in the Serbian texts, according to their semantic classification. Besides, large amount of annotated temporal data may be valuable both for linguists interested in time and language, and, particularly, for computer scientists and mathematicians interested in development of named-entity recognition systems based on machine-learning approaches, whose success is highly dependent on the availability of the extensive training datasets.

Keywords: Temporal Expression Recognition, Temporal Expression Normalization, Finite-State Transducers, Named Entities

1 Introduction

Time has always been a subject of numerous debates within religion, philosophy, and various scientific fields (linguistics, logic, physics, psychology, and history). Considering widely divergent answers to fundamental questions about the nature of time, it is quite difficult to give a clear definition of time applicable to all disciplines. However, despite the existing difficulties in defining or explaining the concept of time, human beings easily manipulate time in the real world. The world is dynamic in its nature and time is fundamental for reasoning about how the world changes. Things that happen and involve change (events) or situations that stay the same for a certain period of time (states) are related by their temporal reference. People use the concept of time to place events or states in sequence one after the other, to establish how long an event or a state lasted, and to specify when an event occurred. Time seems to play the role of an universal reference system that is used to anchor, sequence, measure and compare the intervals occupied by events and states [16].

The way humans process and perceive time is reflected in communication, particularly in linguistic expressions frequently used in everyday speech. The development of systems, tools and devices for electronic communication provided conditions for creating large amounts of natural language text stored in digital format, as well as advanced computational tools for the automatic processing of human language, capable of automatic identification of specific information about temporal entities. The need for these tools arises due to the fact that reasoning about time is critical to effective communication and that most of the information available electronically is time-bounded, in that sense that at different times something may be true and false.

Even though humans manage temporal information very naturally and efficiently during their everyday life, the formalization of temporal referring expressions in the language understood by computers is a challenging and difficult task. First of all, in Serbian, like in any other natural language, the same temporal information can be written in different forms: *13:45 časova* '13:45 am', *1:45 popodne* '1:45 pm', *15 do dva popodne* 'quarter to two in the afternoon' and many others. Furthermore, the use of synonyms such as *sat* and *čas* for a temporal unit 'hour' is also widespread. In addition, Serbian is a highly inflected and a free word order language with a particularly complex number system in which, beside singular and plural, paucal[1] also exists. Since the constructions with numerals require their agreement in gender and number with the nouns they modify, temporal expressions *jedan sat* 'one hour', *dva sata* 'two hours' and *pet sati* 'five hours' use three different inflected forms of a noun *sat* 'hour' - nominative singular, paucal, and genitive plural, respectively. Besides, the numeral *jedan* 'one' behaves as an adjective, also agreeing with the noun in case (e.g. *jednog sata*, *jednom satu*, etc.). Another challenge in formalizing temporal expressions (TEs) arises from the fact that temporal information can often be implicit and linked to an interval which is not given in the text itself but must be taken from context. For instance, events mentioned in a statement do not have to be anchored to precise points in time nor specifically ordered with respect to neighboring events. Therefore, in order to get the correct temporal interpretation it is necessary to rely on semantic content and interpretations or inferences derived from the experience in the form of world knowledge. Since human beings locate events in time with respect to their own reference point relative to the moment of speech, reasoning about the basic temporal entities (times and events) is currently extremely difficult for computers. The "moment of speech" may be more comprehensive reference dependent on the context utterance, which can be seen on the polysemy of time adverbs *sada* or *danas* ('now' or 'today'). Their temporal anchor can be the moment of speaking or writing, but it also has the ability to mark a period much larger than the moment in which the discourse takes place and could be replaced by a somewhat broader notion of temporal anchoring [26].

[1] Paucal is a special kind of grammatical number used with small values (*dva* 'two', *tri* 'three' and *četiri* 'four').

The development of language technology considerably varies from one language community to another. A large number of tools and resources used for the automatic processing of temporal information has been developed mainly for English and for some other languages, such as German, French, Italian, Spanish, Chinese, etc. The previous efforts in the development of Serbian language technologies [11,12,13] provide the necessary basis for the development of the temporal processing system. Since the survival of many European languages in the networked society is questionable and the danger of their extinction in the digital market is greater [27], the need to develop such a system that will contribute to the Serbian language preservation in the digital environment is ever growing. In addition, the existence of such a system can greatly improve the performance of various language technology applications (information extraction, information retrieval, question answering, text summarization, etc.), in general or specific domains. For example, in the field of medicine time appears as one of the core concepts [28,23]. Among other things, during the diagnosis it is very important to know the temporal order of specific symptoms or the length of their duration. It would therefore be useful to develop a system that can easily adapt to new and different requirements and be applied to text from a different time period or domains.

2 The Most Important Existing Resources and Computational Approaches for Temporal Processing

A particular challenge is to properly identify TEs and their values, and to incorporate them into an automatic system capable of extracting temporal meaning from a text [14]. Along with the ever growing amount of electronically available information, temporal processing receives increasing attention in the field of natural language processing over the past twenty years.

The Message Understanding Conferences (MUCs) [5], maintained from 1993-1998 with the aim of evaluating the performance of various existing information extraction systems, have played a significant role by setting a several tasks, such as Named Entity Recognition (NER). Alongside different types of named entities, TEs were also included as targeted classes within the scope of the NER task. The MUC evaluations only covered the recognition of two types of TEs (dates and times) and did not require resolution of their values, while a novel contribution towards the normalization of TEs was made within the scope of the Automatic Content Extraction (ACE) campaigns after 2000 [15]. The first exercise evaluating systems' performance that deals both with recognition and normalization problems was Temporal Expression Recognition and Normalization (TERN) 2004 competition [2], launched by ACE. The TERN task required identification of TEs in free text, as well as normalization of their values, represented in ISO-based format.

2.1 Annotation Schemes

Since temporal information extraction was first included in the context of MUC5 1996, several efforts have been made to define standard ways to annotate the temporal information in texts. At the beginning, TEs were annotated using SGML tag <TIMEX>, characterized by only one attribute - TYPE. Since then, there are a number of annotation schemes, but only three important ones have been extensively used in the past for the development of resources for temporal processing: TIDES, STAG and TimeML.

TIDES Temporal Guidelines [6] have been developed to support research activities under the DARPA TIDES (Translingual Information Detection, Extraction and Summarization) research program. In order to annotate TEs and represent their values according to the ISO-8601 standard format [1], the TIMEX2 annotation scheme emerged. As a very thorough set of guidelines for annotating time expressions, TIDES TIMEX2 extends the TIMEX scheme by widening the range of TEs to be recognized, as well as attributes that specify in greater detail the semantics of a time expression.

STAG (Sheffield Temporal Annotation Guidelines) [24] is the temporal annotation language defined with the aim of identifying events and temporal relations between events or between events and times. The proposed annotation scheme also requires the annotation of time expressions, but it is much less detailed in this regard than the TIDES guidelines, which are more or less completely adopted.

TimeML [22], integrating features of both TIDES and STAG annotation schemes, presents a more general-purpose markup language for TEs, events and temporal relations. In order to provide systematic way for extraction and representation of temporal information as well as to facilitate the exchange of temporal information, TimeML has recently been adopted as an ISO standard ISO-TimeML [3].

2.2 Annotated Corpora

The annotation schemes and the corpora which are annotated according to the schemes together provide objective data resources that can be shared, argued over, and refined by the computational linguistics community. Corpus guided research reveals both the variety and the distribution of the forms of expression in a real sample of language, which is important both for linguists who want to analyze temporal phenomena, and for corpus linguists who employ the annotated data in training and evaluating algorithms for automatic temporal processing.

The existing annotated resources that are most widely employed by the researchers studying different temporal phenomena were developed for each annotation standard, providing an objective basis to evaluate competing algorithms. For the evaluation of automatic TIMEX2 annotation performance in the TERN 2004 competition, the TERN corpus was employed, including both English and Chinese data. Even though the TERN corpus is the most reliably annotated resource for temporal processing developed so far, it restricts the temporal analysis to identification and normalization of TEs. TimeBank is the human-annotated corpus marked up for TEs, events and temporal relations with respect to the most recent annotation language TimeML. Based on TimeBank, the TempEval corpus was developed in the scope of the TempEval evaluation exercise with the aim of simplifying the identification of temporal relations.

2.3 Computational Approaches for Identification and Normalization of Temporal Expressions

There are several computational approaches taken towards the identification and normalization of TEs in natural text. As for any other information extraction task, there are two basic approaches for the recognition of TEs: rule-based (knowledge-based) and machine-learning (statistical or data-driven) methods. Systems based on hand-crafted rules achieve very high accuracy [18], but good response requires significant efforts of linguists involved in policy development. It has been shown that for the TEs identification task only both techniques can be successfully employed as long as sufficient training data is available for data-driven systems. However, when it comes to TE normalization, rule-based approaches are by far more appropriate than data-driven techniques that cannot solve this problem anywhere near as successful methods based on rules [19,8,25,4]. Data-driven and hybrid systems were more popular among those approaches that tackled event processing, and show better results than rule-based systems in this task [10,9].

3 System for Recognition and Normalization of Temporal Expressions in Serbian

Our system for automatic recognition and normalization of TEs is a part of the system for named entity recognition and normalization in Serbian texts. The TE system developed for Serbian is a rule-based system that relies on lexical resources and handles both absolute time (e.g. *17. marta 2001. godine* 'March 17^{th} 2001', *19:25 časova* '19:25 o'clock', *leta 2000.* 'summer 2000', *pet nedelja* 'five weeks') and relative time (e.g. *15. aprila* 'April 15^{th}', *sutra uveče* 'tomorrow evening', *sledeće godine* 'next year', *nekoliko dana* 'a few days', *par nedelja* 'a couple of weeks'). Its role is recognizing temporal expressions in unstructured texts and re-interpreting their temporal semantics in a standard format, according to the TimeML annotation guidelines, specified in [3,21]. The recognition of events and temporal relations which may exist among them remains beyond the scope of the current version of the system.

Earlier version of our system, based on a large-coverage set of finite state transducers (FSTs), was designed to perform both recognition and normalization of TEs as a single stage task. Even though the evaluation results were quite good, showing the ability of the system to correctly assign normalized values for all correctly recognized expressions, further development was hindered with the ever growing collection of transducers. In order to find a solution that best resolves the issue of the increasing complexity, recognition and normalization tasks are done separately and the collection of used transducers is transformed into the cascade of transducers.

3.1 Recognition of Temporal Expressions

Temporal expressions are natural language phrases that give information about when something happened, how long something lasted, or how often something occurred. They are usually detected in the input text by the presence of lexical triggers - reserved words that indicate the concept of time, date and duration. Our system takes into account triggers of the following kind:

- nouns (e.g. *sat* 'hour', *vikend* 'weekend', *dan* 'day', *godina* 'year', *vek* 'century', *podne* 'noon', *petak* 'Friday', *septembar* 'September', *noć* 'night', *jutro* 'morning', *popodne* 'afternoon');
- specialized time patterns (e.g. 16:45, 27.05.2006., 11/30/2005, 1998, *1970-tih* '1970s');
- adjectives (e.g. *prošli* 'past', *tekući* 'current', *sledeći* 'next', *devedesetih* 'nineties');
- adverbs (e.g. *mesečno* 'monthly', *dnevno* 'daily', *nedeljno* 'weekly', *večeras* 'tonight', *danas* 'today', *juče* 'yesterday');
- numerals (e.g. 2 (as in 'Ivan arrived at 2'), *pet* 'five', 3^{rd} (as in 'She arrived on the 3^{rd}')).

The lexical context surrounding detected triggers represents relevant information required for the correct full extent determination and later normalization process. Therefore, some of the non-trigger words used as modifiers (e.g. *početkom* 'in the early', *manje od* 'less than'), as well as expressions denoting sets of times (e.g. *svakog* 'every', *dva puta* 'twice') are included in the expression extent, while prepositions are never marked and cannot represent a part of a temporal expression (e.g. *tokom* ***marta*** 'during March', *nakon* ***20 dana*** 'after 20 days').

3.2 The Cascade of Finite-State Transducers

Resources for natural language processing of Serbian are being developed using the finite-state methodology as introduced by Maurice Gross and LADL (*Laboratoire d'Automatique Documentaire et Linguistique*) laboratory [7]. For development and application of these resources the Unitex corpus processing system was used [20]. The processing of TEs is carried out on a text having undergone a part-of-speech tagging and morphological analysis. On the input text general-purpose lexical resources (electronic dictionaries and dictionary finite-state transducers) are applied, and their role is text tagging, with lemmas, grammatical categories and semantic features. After a text is being tagged this way, the system for TE recognition is applied.

The system for recognition of TEs developed for Serbian is based on a cascade of FSTs - CasSys [17], integrated in the Unitex corpus processor. Our FST cascade currently consists of 14 finite transductions and their role is to determine the full span and type of each detected expression, in accordance with the TimeML schema (DATE, TIME, DURATION, and SET). Therefore, the graphs in this large-coverage grammar are designed to recognize expressions denoting calendar dates (1), times of day (2), durations (3) and sets of recurring times (4).

(1)		
13. juna 2008. godine	‘June 13th 2008’	+time+date+abs
proleća 2000.	‘spring 2000’	+time+date+abs
8. aprila prošle godine	‘April 8th last year’	+time+date+rel
desetog jula	‘10th July’	+time+date+rel
petak	‘Friday’	+time+date+rel
sutra	‘tomorrow’	+time+date+rel
(2)		
19:35 h	‘19:35’	+time+hour+abs
5 i 25 popodne	‘5 and 25 in the afternoon’	+time+hour+abs
jutros	‘this morning’	+time+hour+rel
kasno sinoć	‘late last night’	+time+hour+rel
(3)		
hiljadu i 700 godina	‘one thousand and 700 years’	+time+duration+abs
narednih mesec i po dana	‘next month and a half’	+time+duration+abs
nekoliko dana	‘a few days’	+time+duration+rel
(4)		
svake godine	‘every year’	+time+set
svakog decembra	‘every December’	+time+set
tri puta nedeljno	‘three times a week’	+time+set
dva meseca svake godine	‘two months every year’	+time+set

Each transduction is defined by a set of patterns. For the most frequent variant forms of dates and times represented in Serbian, corresponding FSTs were built and applied to text to recognize patterns described in the input alphabet. When the pattern was matched, the output alphabet specified the action to be taken. For instance, FST *Datum* in Fig. 1 recognizes some possible date patterns that consist of a day (written using digits or letters) followed by month (written in letters or Roman digits) followed by year (written using digits), as well as incomplete date expressions in which year is omitted. The output contains the TE described in the input and with addition of a lexical tag that can be used in subsequent FSTs. Semantic markers associated to recognized expressions provide useful information primarily regarding the type of the named entity (+time), as well as temporal expression (+date, +hour[2], +duration, +set), as given in (1-4). Additional information concerning the type of TE is provided by semantic markers +abs and +rel, used to indicate absolute and relative TEs, respectively.

[2] Corresponds to TimeML type TIME.

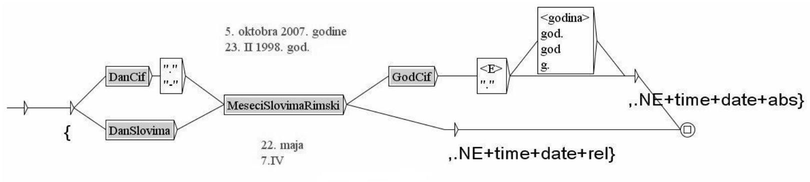

Fig. 1: One path from FST *Datum* that recognizes complete and incomplete date expressions.

Applied to a text in a predefined order, FSTs for temporal expressions first track down the most certain patterns of dates, times and durations that can be retrieved with a high degree of certainty, while the retrieval of others is postponed. The produced lexical tags, used in subsequent FSTs, enable detection of more complex expressions, such as temporal ranges (5) and conjoined expressions (6) or combinations of calendar dates and times-of-day (7).

(5)
od {{8. marta,.NE+time+date+rel} do {7.aprila,.NE+time+date+rel},.NE +time+date+period} 'from March 8^{th} to April 7^{th}'

(6)
izmedju {{18 i 30,.NE+time+hour+abs} i {19 časova,.NE+time+hour+abs},. NE+time+hour+period} 'between 18:30 and 19 pm'

(7)
{{15. marta,.NE+time+date+rel} {oko 2 sata,.NE+time+hour+abs},.NE +time+hour} 'March 15^{th} around 2 o'clock'

Cascade is a simple and effective way of organizing FSTs that may greatly increase precision and speed of the system, as well as containment of ambiguity. For example, system's ability to manage priorities between patterns disambiguates the meaning of TEs based on the context information, distinguishing whether an ambiguous unit phrase refers to a point or duration. The expression *12 časova* '12 o'clock/12 hours' is used to indicate time of the day or duration of time. For both entities - hour and duration of time, several rules are built and, when applied in a particular order, the precision of recognition is increased. In the sequence *od 12 časova* the time expression is recognized as representing duration, if preceded by *u trajanju* 'lasting' or some other trigger word; the subsequent FST in a cascade can then safely recognize the other occurences of the same sequence as a time of a day. Furthermore, there are a lot of appearances of numerals which do not necessary have to refer to time of a day, and the cascade helps us providing the right context for disambiguation. For instance, numerals that occur together with some already tagged dates could be reliable indicators of some time patterns after which words *čas* or *sat* 'hour' do not appear (Fig. 2).

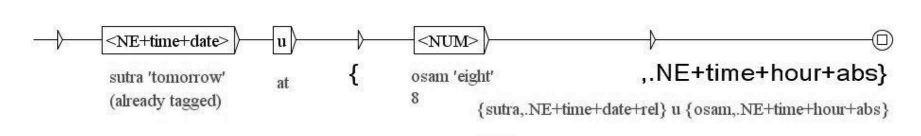

Fig. 2: A simplified path in a FST that a numeral occuring after already marked date tags as time.

Lexical tags produced by FSTs, even though the most convenient for the use of subsequent FSTs in the cascade, are not useful for other applications and at the end are converted to the XML tags (8).

(8)
{*hiljadu i 700 godina*,.NE+time+duration+abs}
<time.duration.abs>*hiljadu i 700 godina*</time.duration.abs>
'one thousand and 700 years'

3.3 Normalization of Temporal Expressions

For the normalization of recognized TEs we constructed large collection of Unitex finite-state transducers. For instance, the FST given in Fig. 3 illustrates the way one of the usual representations of the time of a day, consisting of an hour (written using digits or letters) that is optionally followed by minutes (written in letters), is normalized. The fact that the hour is preceded by an adverb *popodne*

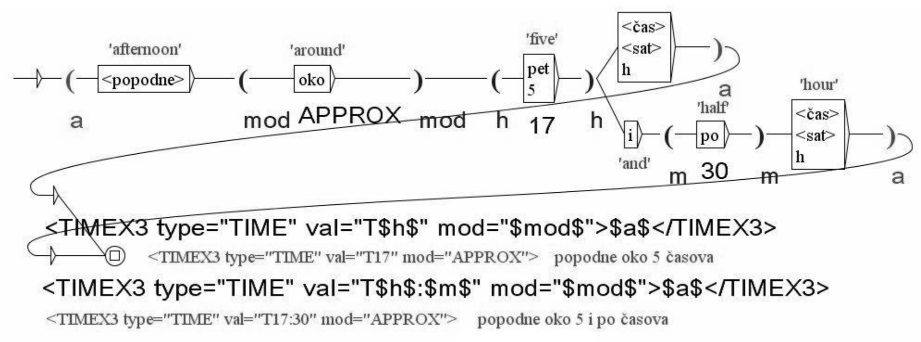

Fig. 3: An illustrative FST that shows how two cases of absolute temporal expressions related to a time of a day are normalized.

'afternoon' affects its value specified in the output (17:30 instead of 5:30). Those two recognized units (hours and minutes) become the values of the variables \$h\$ and \$m\$, respectively, that will be used in the output to produce a normalized form, as values of a tag attribute **value**.

The normalization of multi-word numerals that often appear in duration expressions (the first line in example (3)) is done by the dictionary FSTs. The output of these transducers is a lemma and the values of the grammatical categories of a recognized form, given in the same format that is used for e-dictionaries (9).

(9)
hiljadu i 700 godina,1700.NUM+C+v5 ‘one thousand and 700 years’

Their lemmas could be retrieved from those applied dictionaries and used in the output tag as values of a tag attribute **value**. The required output is the recognized TE embedded in XML tags with appropriately assigned values of attributes, according to the TimeML.

Each detected TE was marked up with the <TIMEX3> tag, which may contain the following attributes: **type**, **value**, **temporalFunction**, **valueFromFunction**, **mod**, **quant** and **freq**.

Non-optional attribute **type**, provided in the cascade output, may have the following values: DATE, TIME, DURATION, and SET, as explained in 3.1.

The attribute **value** contains the normalized form of the detected TE that is derived from the ISO 8601 standard format for representing time values [1].

Absolute TEs are translated in the normalized form by means of simple regular expressions, as explained in Fig. 3, while relative TEs require additional information represented by the values of the following two attributes. Attribute **temporalFunction** is a binary attribute which indicates whether or not further resolution of the TE is needed. Relevant for the value calculation purposes, attribute **valueFromFunction** consists of the operator (“+”, “-”, or “=”) that has to be applied to calculate the final value, and the quantity (n≥0) that has to be added or subtracted to the selected temporal anchor. For instance, given the example (10), referenced expression will be modified with respect to the number (2), magnitude (years) and temporal direction (ago). Thus, the final value will be calculated by subtracting a number of years for the value of the referenced timex, e.g. document’s creation date or the nearest previous absolute time expression.

(10)
dve godine ranije ‘two years ago’
<TIMEX3 type=“DATE” val=“XXXX-XX-XX” temporalFunction=“true” valueFromFunction=“-2D”>

Assigning the correct value of TE is crucial since it is used for further analysis of the documents. The semantics of temporal expressions that are modified in some way is expressed through the assigned value of the optional **mod** attribute (illustrated in Fig. 3 and example (11)). Possible values for **mod** used at this moment are illustrated in examples (11).

(11)
početkom 1999. 'early1999'
<TIMEX3 type="DATE" val="1999" mod="START">
sredinom marta 'Mid-March'
<TIMEX3 type="DATE" val="XXXX-03" mod="MID">
krajem zime 2007. 'late winter'
<TIMEX3 type="DATE" val="2007-WI" mod="END">
oko 6 sati ujutru 'around 8 o'clock'
<TIMEX3 type="TIME" val="T08:00" mod="APPROX">
oko hiljadu i 700 godina 'about one thousand and 700 years'
<TIMEX3 type="DURATION" val="P1700Y" mod="APPROX">
više od 15 dana 'more than 15 days'
<TIMEX3 type="DURATION" val="P15D" mod="MORE_THAN">
gotovo dva meseca 'nearly two months'
<TIMEX3 type="DATE" val="P2M" mod="LESS_THAN">

Values of the **mod** attribute START, MID and END capture the basic semantics of lexicalized aspect markers (early, start, mid), while APPROX, MORE_THAN and LESS_THAN capture the basic semantics of quantifier modifiers (approximately, no more than). Attributes **quant** and **freq** are used in addition to the value of sets of recurring times. Attribute **quant** is generally a literal from the text that quantifies over the expression, while the attribute **freq** contains an integer value and a time granularity that represent the frequency within the set (12).

(12)
svake godine 'every year'
<TIMEX3 type="SET" val="P1Y" quant="EVERY">
tri puta nedeljno 'three times a week'
<TIMEX3 type="SET" val="P1W" freq="3X">

4 Conclusions and Future Work

In this paper we presented the system for recognition and normalization of TEs in Serbian texts, which is the most complete approach that is currently available for Serbian language. It has been shown that the system for recognition and normalization of TEs in Serbian natural language texts, based on a finite-state transducers methodology, is effective and competitive with respect to other techniques. After implementation of the cascaded FSTs, re-evaluation needs to be done in order to verify the improvement of the system's overall performance.

We have produced the first versions of temporally annotated corpora that could be useful to the research community interested in both analysing different temporal phenomena and producing a machine-learning NER system for Serbian.

Future research in temporal processing is needed to complete the tagger, in particular for recognition of events and temporal relations that hold between temporal entities. We also plan to test our work when applied to various domains.

References

1. *ISO: ISO 8601 Data Elements and Interchange Formats – Information interchange – Representation of Dates and Times.* Geneva, Switzerland, 2004.
2. *The TERN 2004 Evaluation Plan (draft):Time Expression Recognition and Normalization. Technical report.* McLean, Virginia, 2004.
3. *ISO: ISO/DIS 24617-1 Language Resources Management – Semantic Annotation Framework (SemAF) – Part 1: Time and Events (SemAF-Time, ISO-TimeML).* Geneva, Switzerland, 2009.
4. Nathanael Chambers. Navytime: Event and Time Ordering From Raw Text. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 73–77, Atlanta, Georgia, USA, 2013. Association for Computational Linguistics.
5. Nancy Chinchor. MUC-7 Named Entity Task Definition (version 3.5). In *Proceedings of the 7th Message Understanding Conference (MUC-7)*, Fairfax, VA, 1998.
6. Lisa Ferro, Beth Sundheim, and George Wilson. TIDES Temporal Annotation Guidelines – Draft Version 1.0. Technical Report Technical report, The MITRE Corporation, McLean, Virginia, 2000.
7. Maurice Gross. Local Grammars and their Representation by Finite Automata. In M. Hoey, editor, *Data, Description, Discourse. Papers on the English Language in Honour of John McH Sinclair*, pages 26–38, 1993.
8. Valentin Jijkoun, Mahboob Alam Khalid, Maarten Marx, and Maarten De Rijke. Named Entity Normalization in User Generated Content. In *Proceedings of SIGIR 2008 Workshop on Analytics for Noisy Unstructured Text Data, AND'08*, pages 23–30, 2008.
9. Hyuckchul Jung and Amanda Stent. ATT1: Temporal Annotation Using Big Windows and Rich Syntactic and Semantic Features. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 20–24, Atlanta, Georgia, USA, 2013. Association for Computational Linguistics.
10. Oleksandr Kolomiyets and Marie-Francine Moens. KUL: A Data-Driven Approach to Temporal Parsing of Documents. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 83–87, Atlanta, Georgia, USA, 2013. Association for Computational Linguistics.
11. Cvetana Krstev. *Processing of Serbian – Automata, Text and Electronic Dictionaries.* University of Belgrade, Faculty of Philology, Belgrade, 2008.
12. Cvetana Krstev, Ivan Obradović, Miloš Utvić, and Duško Vitas. A System for Named Entity Recognition Based on Local Grammars. *Journal of Logic and Computation*, 24(2):473–489, April 2014.
13. Emeline Lecuit, Denis Maurel, Duško Vitas, and Cvetana Krstev. Temporal Expressions: Comparisons in a Multilingual Corpus. In Zygmunt Vetulani, editor, *Proceedings of the 4th Language and Technology Conference: Human Language*

Technologies as a Challenge for Computer Science and Linguistics, November 6-8, 2009, Poznań, Poland, Poznań, Poland, 2009. Widawnictwa Elektroniczne S.A.
14. Inderjeet Mani, James Pustejovsky, and Robert Gaizauskas. *The Language Of Time: A Reader.* Oxford University Press, New York, 2005.
15. Inderjeet Mani and George Wilson. Robust Temporal Processing of News. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 69–76, Stroudsburg, PA, USA, 2000. Association for Computational Linguistics.
16. Georgiana Marşic. *Temporal Processing of News: Annotation of Temporal Expressions, Verbal Events and Temporal Relations.* PhD thesis, Wolverhampton, UK, December 2011.
17. Denis Maurel, Nathalie Friburger, Jean-Yves Antoine, Iris Eshkol-Taravella, and Damien Nouvel. Transducer Cascades Surrounding the Recognition of Named Entities. *Traitement Automatique des Langues*, 52(1):69–96, 2011.
18. Andrei Mikheev, Claire Grover, and Marc Moens. Description of the LTG System Used for MUC-7. In *Proceedings of the 7th Message Understanding Conference (MUC-7).* Fairfax, VA, 1998.
19. Matteo Negri and Luca Marseglia. Recognition and Normalization of Time Expressions: ITC-irst at TERN 2004. Technical report. Technical report, Information Society Technologies, 2005.
20. Sébastien Paumier. *Unitex 3.0 User Manual*, 2011. Available at `http://www-igm.univ-mlv.fr/~unitex/UnitexManual3.0.pdf`.
21. James Pustejovsky, Harry Bunt, Kiyong Lee, and Laurent Romary. ISO-TimeML: an International Standard for Semantic Annotation. In *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, pages 394–397, Paris, France, 2010. ELRA.
22. James Pustejovsky, José M Castano, Robert Ingria, Roser Sauri, Robert J Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R Radev. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *New directions in question answering*, pages 28–34, 2003.
23. Ruth M Reeves, Ferdo R Ong, Michael E Matheny, Joshua C Denny, Dominik Aronsky, Glenn T Gobbel, Diane Montella, Theodore Speroff, and Steven H Brown. Detecting Temporal Expressions in Medical Narratives. *International Journal of Medical Informatics*, 82(2):118–127, 2013.
24. Andrea Setzer. *Temporal Information in Newswire Articles: An Annotation Scheme and Corpus Study.* PhD thesis, University of Sheffield.
25. Jannik Strötgen and Michael Gertz. HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation, ACL 2010, 15-16 July 2010*, pages 321–324, Uppsala, Sweden, 2010. Association for Computational Linguistics.
26. Ilijana Čutura. O jednom semantičkom tipu vremenskih priloga. *Nasledje*, 3(4):117–129, 2006.
27. Duško Vitas, Ljubomir Popović, Cvetana Krstev, Ivan Obradović, Gordana Pavlović-Lažetić, and Mladen Stanojević. *The Serbian Language in the Digital Age.* META-NET White Paper Series. Georg Rehm and Hans Uszkoreit (Series Editors). Springer, 2012.
28. Li Zhou and George Hripcsak. Temporal Reasoning with Medical Data – A Review with Emphasis on Medical Natural Language Processing. *Journal of Biomedical Informatics*, 40(2):183–202, 2007.