Semi-automatic Annotation of the Romanian TimeBank 1.2

Corina Forăscu
Faculty of Computer Science,
Al.I. Cuza University of Iaşi, Romania;
Research Institute for Artificial Intelligence,
Romanian Academy, Bucharest
16, Gen. Berthelot, Iaşi - 700483
corinfor@info.uaic.ro

Radu Ion
Research Institute for Artificial Intelligence,
Romanian Academy, Bucharest
13, Calea 13 Septembrie,
radu@racai.ro

Dan Tufiş
Research Institute for Artificial Intelligence,
Romanian Academy, Bucharest;
Faculty of Computer Science,
Al.I. Cuza University of Iaşi, Romania
13, Calea 13 Septembrie,
tufis@racai.ro

Abstract

Temporal information is essential in many Natural Language Processing applications. The paper presents our activities and research towards obtaining a parallel version, English-Romanian, of the TimeBank 1.2. corpus, annotated according to the TimeML 1.2.1. standard. The automatic import from English to Romanian can be done automatically for 96.53% of the temporal markups; it will determine the applicability of the general temporal theories when used in Romanian. The parallel corpus will be at the basis of other temporal processing tools.

Keywords

Temporal information, language processing, lexical alignment, annotation import.

1. Introduction

The temporal information can be verbalized in natural language by:

- Time-denoting temporal expressions – references to a calendar or clock system, expressed by NPs, PPs, or AdvPs, as in *Monday*; *tomorrow*; *the next week*.
- Event-denoting temporal expressions – explicit/implicit/vague references to an event; syntactically they are realized by:
  - sentences – more precisely their syntactic head, the main verb, as in *John listens to the music.*
  - noun phrases, as in *Israel will delay a military strike against Iraq*.
  - adjectives, predicative clauses or prepositional phrases.

Recent work in document analysis started focusing on the temporal information in documents, mainly for their use in many practical Natural Language Processing (NLP) applications such as:

- lexicon induction, translation and linguistic investigation, using very large annotated corpora;
- question answering (questions like "when", "how often" or "how long");
- information extraction or information retrieval;
- machine translation (translated and normalized temporal references; mappings between different behavior of tenses from language to language);
- discourse processing: temporal structure of discourse and summarization (temporally ordered information, biographic summaries)

To join this growing interest for temporal information, we have decided to study the events, time expressions, their temporally anchoring and ordering when applying general temporal theories to the Romanian language.

In order to have linguistic evidence of how temporal information is really used in Romanian, as source of evidence to inform and substantiate the theory, we have chosen to use the TimeML 1.2. annotation standard (Saurí et al., 2006) together with the TimeBank 1.2. corpus, an English news corpus manually annotated and widely used in the temporal community (Pustejovsky et al., 2006). The manual temporal annotation is very time consuming, expensive (Pustejovsky et al., 2002) and error-prone, hence it would be useful to use some help or back-up from the same annotation applied to a parallel text.

While there are many Romanian and parallel linguistic resources (Cristea, Forăscu, 2006), none supports temporal annotation and currently, to our knowledge, there are no software tools to deal with all temporal information in Romanian texts.

Section 2 gives a brief state of the art in the field of temporal annotations and information in NL. In section 3 the TimeML standard and the TimeBank corpus are briefly presented. The next section details the work toward obtaining the parallel English-Romanian corpus, the pre-processing, alignments and annotation import performed on it. As the paper presents "work-in-progress", the last section presents the conclusions and discusses future plans with regard to the corpus in order to see how temporal linguistic theories can be applied to Romanian, and applications to be developed by using it.

2. Temporal Information in NL

Much of the early work has adopted Allen's 13 temporal binary relations between time intervals (Allen, 1984), and has used meaning representations augmented with temporal variables: in (Reichenbach, 1947), the verb tenses are classified according to the ordering of three parameters: the points of speech, of the event and of reference.

The work in the temporal annotation area has been fostered by the Message Understanding Conference, MUC-7[1] of 1998. Main activities connected to temporal information and different types of temporal annotation schemes have been developed since then (Mani et al., 2005). The most used annotation schemes are TIMEX2 (Ferro et al., 2005) and TimeML (Saurí et al., 2006). TIMEX2 is a component technology in ACE[2], conceived to fill the temporal attributes for extracted relations and events. TimeML is more complex and it treats unitarily the temporal aspects of texts, hence it is useful in much more applications.

The TimeML standard integrates together two annotation schemes: TIMEX2 and Sheffield STAG (Setzer 2001 – as a first complete mention of STAG[3]), a fine-grained annotation scheme capturing events, times and temporal relations between them, as well as other emerging work (Katz and Arosio 2001).

Time corpora and taggers have been created mainly for English, but French, German, Spanish, Chinese, Arabic and Korean[4] start to become prominent languages in the field.

Regarding the Semantic Web, some significant efforts have been invested in order to develop ontologies of time, for expressing the temporal content of web pages and temporal properties of web pages and web services (SUMO[5], CYC[6] among others). DAML-Time ontology (Hobbs, 2002) is a collaborative effort towards standardizing the basic topological temporal relations on instants and intervals, measures of duration, clock and calendar units, months and years, time and duration stamps, including temporal aggregates (*for the last four years*), deictic time (*now*) and vague temporal concepts *(recently, soon*).

TIMEX2 scheme is compatible with the KSL-Time ontology[7], while TimeML is mapped to the DAML-Time Ontology (Hobbs and Pustejovsky, 2003), hence advanced inferential capabilities based on information extracted from text are better supported.

3. TimeML 1.2.1. and TimeBank 1.2.

The TimeML standard has been developed for the purpose of automatically extracting information about the event-structure of narrative texts, and has been applied mainly to English news data. The mark-up language consists of a collection of tags intended to explicitly outline the information about the events reported in a given text, as well as about their temporal relations.

The TimeML metadata standard marks:

- Events through the tags:
  - `EVENT`: it indicates situations that happen or occur, states or circumstances in which something obtains or holds true: *We are waiting for him.*
  - `MAKEINSTANCE`: it marks how many different instances or realizations have a given event; the tag also carries the `tense` and `aspect` of the verb-denoted event: *He learns* ***twice*** *today.*
- temporal anchoring of events through the tags:
  - `TIMEX3`: it marks: times of a day, dates – calendar dates or ranges, durations: *15 November*; *two days*.
  - `SIGNAL`: it marks function words that indicate how temporal objects are to be related to each other.
- links between events and/or timexes through the tags:
  - `TLINK` – Temporal Link – indicates 13 types of temporal relations between two temporal elements (event-event, event-timex).
  - `ALINK` – Aspectual Link (of type `Initiation`, `Culmination`, `Termination`, `Continuation`) – marks the relationship between an aspectual event and its argument event.
  - `SLINK` – Subordination Link (of type `Modal`, `Factive`, `Evidential`, `Negative`) – marks contexts introducing relations between two events.

The creation of the TimeBank corpus started in 2002 during the TERQAS[8] workshop, and it should be considered preliminary. In (Boguraev, Ando, 2006) it is shown that the corpus still needs improvements and reviews. The dimension of the corpus (4715 sentences with 10586 unique lexical units, from a total of 61042 lexical units) might be too small for robust statistical learning and the annotation inconsistencies (incomplete or inconsistent temporal or subordination links, perfectible event classification, incomplete annotation of the tense and aspect for some event) require corrections. Now it consists of 183 news report documents, with XML markups for document format and structure information, and named entity recognition (`ENAMEX, NUMEX, CARDINAL` from MUC-7), sentence boundary information. TimeBank 1.2. is temporally annotated according to the TimeML 1.2.1 standard. Some statistics are shown in table 1.

[1] http://www.itl.nist.gov/iaui/894.02/related_projects/muc/
[2] http://www.nist.gov/speech/tests/ace/index.htm
[3] We are grateful to our reviewers for pointing this
[4] http://complingone.georgetown.edu/~linguist/
[5] http://ontology.teknowledge.com/rsigma/arch.html#Temporal
[6] http://www.cyc.com/cycdoc/vocab/time-vocab.html
[7] http://www.ksl.stanford.edu/ontologies/time/

[8] http://www.timeml.org/site/terqas/index.html

Table 1. Statistics on English TimeBank 1.2

TimeML tags	#
events	7935
instances	7940
timexes	1414
signals	688
alinks	265
slinks	2932
tlinks	6418
TOTAL	**27592**

The corpus is distributed through LDC[9] (Pustejovsky et al., 2006) and it can be browsed online[10].

4. Towards an English-Romanian temporal annotated corpus

In the following subsections we detail the main steps towards the targeted corpus (translation, preprocessing, alignment, and annotation import) and comment on the encountered problems and their (possible) solutions.

4.1 Corpus translation

In order to obtain the Romanian version of the raw text in TimeBank, the corpus was distributed to two Master students in Computational Linguistics with strong background in English and Romanian philology and translation[11].

As the next step is the alignment of the English and Romanian versions of the corpus, a minimal set of translation recommendations was elaborated, in order not only to ensure a literal translation - one which keeps as close as possible to the original version, but also to permit a best-possible word-alignment process. Some basic translation principles are the followings:

- The sentences are translated in a 1:1 correspondence, whenever the language permits it, so that the sentence-alignment is directly obtained through translation.
- The translation equivalents have as much as possible the same part-of-speech; when the English word has a Romanian cognate (*manually – manual*), this is used in translation, and not its Romanian paraphrase (*de/cu/la mână*).
- All words are translated and stylistic variations are avoided, so as not to introduce words or expressions without an English equivalent.
- The tense of verbs is mapped onto its corresponding Romanian one, the modifications being accepted only on linguistic grounds, but not stylistic.
- The format of the dates, moments of day and numbers conforms to the norms of written Romanian.

A manual check performed on the parallel corpus allowed us to detect and correct some lacks and inconsistencies in the way the translators treated especially the header of the documents.

In the 4715 sentences (translation units) of the current version of the corpus there are 65375 lexical tokens, including punctuation marks, representing 12640 lexical types.

4.2 Preprocessing the parallel corpus

In order to run the lexical aligner, the English and Romanian raw texts have to be preprocessed so as to obtain the corpus in the required format. Thus, the texts are tokenized, POS-tagged, lemmatized and chunked using the TTL[12] module (Ion, 2007). This module assembles the bitext in an XML format similar to the XCES one (Ide et al., 2000). Following, there is a brief description of the preprocessing operations:

- tokenization: closely follows the MtSeg model (Armstrong, 1996) dealing with multi-word expressions and clitic splitting (*arătat-<u>o</u> – showed <u>it</u>*) by using lists of those for every language;
- POS tagging: implements the TnT POS tagger (Brants, 2000) enriching it with some heuristics to determine the part-of-speech of an unknown word (only open-class words are considered because the grammatical categories of functional words are thought to be known for a given language);
- lemmatization: a stochastic process which automatically learns lemmatization rules from a lexicon containing triples word form, lemma and POS tag;
- chunking: non-recursive chunks are recognised using a set of regular expressions defined over sequences of POS tags: noun phrases, adjectival phrases (*the most beautiful – cea mai frumoasă*), adverbial phrases, prepositional phrases (*by David – de David*) and verb complexes (*will stay – vor sta*).

4.3 Lexical alignment

Because our COWAL combined word alignment software (Tufiş et al., 2006) is currently under major optimization, for the work reported here only YAWA, one of the two word aligners of the COWAL, was used. YAWA is a four stage lexical aligner[13] that uses bilingual translation lexicons (Tufiş, 2002) and phrase boundaries detection to

[9] http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T08
[10] http://www.timeml.org/site/timebank/browser_1.2/index.php
[11] The reviewers' suggestion – to use professional translators – was our first intention, but impossible to put in practice

[12] **T**okenizing, **T**agging and **L**emmatizing free running texts
[13] Currently, YAWA only supports Romanian to English lexical alignment.

align words of a given bitext. In each of the first three stages, YAWA adds new links to those already created in the previous steps, without deleting from the existing ones. Here is a short description of the alignment stages (all evaluation scores are computed over the data in the Shared Task on Word Alignment, Romanian-English track organized at the ACL2005 (Martin, et al. 2005)):

1. Content words alignment: based on the translation lexicons, the nouns, verbs, adjective and adverbs are aligned. After this stage, YAWA has a high precision, but the recall is improved during the next steps: P = 94.08%, R = 34.99%, F = 51.00%.
2. Inside-Chunks alignment: after a chunk-to-chunk matching based on the first stage, YAWA uses simple empirical rules to align the words within the corresponding chunks; for example a Romanian noun aligned to an English one preceded by an English determiner will be also linked to the determiner (*the girl – fata*). The evaluation after this step gives P = 89.90%, R = 53.90%, F = 67.40%.
3. Alignment in contiguous sequences of unaligned words: using the POS-affinities of these unaligned words and their relative positions, YAWA attempts to heuristically match them.
4. Correction phase: the wrong links introduced mainly in stage 3 are now removed.

The current evaluation of YAWA (P = 88.80%, R = 74.83%, F = 81.22%) (Tufiş et al., 2006) shows a significant improvement over the accuracy reported in (Martin, et al. 2005). The COWAL combiner of YAWA and MEBA word aligners was rated the best out of 37 systems participating in the Shared Task, with the following evaluation scores: P = 87.17%, R = 70.25%, F = 77.80% (Tufiş et al., 2005). As YAWA has already achieved a very good accuracy it can be successfully used on its own[14].

The automatic alignment performed on the TimeBank parallel corpus produced 91714 alignments out of which 25346 are NULL-alignments. In order to obtain an optimal transfer of the temporal annotations from the English version onto the Romanian one, all the alignments were manually checked using MTKit (Ceauşu, 2005). We found that most of the wrong alignments are due to incorrect tokenization of some numbers and values, incorrect POS-tagging mainly for Romanian possessive pronouns, English negations (*no*, *n't*, *neither*) and English adjectives (*lower*, *higher*, *smallest*).

[14] We estimate that when the new COWAL platform will be finalized, the expected F-measure score will be higher than 85%.

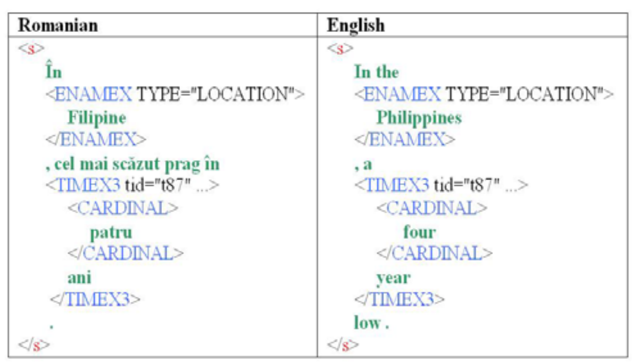
4.4 Import of the temporal mark-up

The translation of the English part of the TimeBank corpus followed the sentence XML structure so it was possible to parse the English corpus and for every sentence XML tag, to extract its content and replace it with the Romanian translation. Within a sentence, due to the nature of the Romanian translations, we can not assume that the word ordering in English is completely preserved into Romanian and also that English received a literal (almost word by word) translation into Romanian. Thus, we need to use the Romanian to English lexical alignment to transfer the XML markup from English to Romanian because, otherwise, we could obtain the Romanian translation in a shuffled form if the word order was not preserved. The transfer algorithm goes as follows:

- for every pair of sentences Sro and Sen from the TimeBank parallel corpus with the Ten English equivalent sentence (Ten is the same sentence – same raw text – as Sen with the exception that Ten has the XML structure that we want to transfer).
- construct a list **E** of pairs of English text fragments with sequences of English indexes from $\mathbf{S}_{en}$ and $\mathbf{T}_{en}$. Due to the fact that the tokenization of $\mathbf{S}_{en}$ is different from that of $\mathbf{T}_{en}$, the list **E** is needed in order to map English text fragments from $\mathbf{T}_{en}$ with sequences of indexes from $\mathbf{S}_{en}$ so as to be able to use the Romanian lexical alignments which exist relative to these indexes. For instance, looking at Figures 1 and 2,

  E = {<"In the"; 1,2>, <"Philippines"; 3>, <", a"; 4,5>, <"four"; 6>, <"year"; 7>, <"low ."; 8,9>}.

Figure 1 An example of the XML markup transfer from English to Romanian



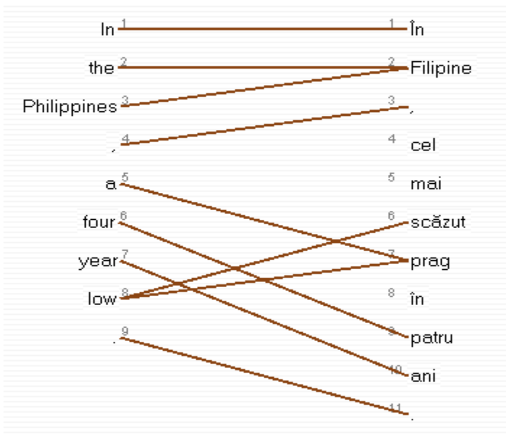
- add to every element of **E** the XML context in which that text fragment appeared.

  For instance, the first element of **E**, <"In the"; 1,2>, appears in the s context and 4th element, <"four"; 6> appears in 3 contexts: s, TIMEX3, and CARDINAL. Thus the list **E** becomes

{<"In the"; 1,2; s>, <"Philippines"; 3; s,ENAMEX>, …} (for every tag, its attributes - if present - are stored).

- construct the list **RW** of Romanian words along with

Figure 2 The lexical alignment for the sentences in Figure 1



the transferred XML contexts using **E** and the lexical alignment between $\mathbf{S}_{ro}$ and $\mathbf{S}_{en}$. If a word in $\mathbf{S}_{ro}$ is not aligned, the top context for it, namely s, is considered.

Using the example in Figures 1 and 2, **RW** = {<"În"; s>, <"Filipine"; s,ENAMEX>, …}.

- construct the final list **R** of Romanian text fragments from **RW** by conflating adjacent elements of **RW** that appear in the same XML context. Output the list in XML format (Figure 1 - the result of XML markup transfer).

Table 2 Statistics on the Romanian TimeBank 1.2

TimeML tags	#	% transfered
events	7703	97.07
instances	7706	97.05
timexes	1356	95.89
signals	668	97.09
alinks	249	93.96
slinks	2831	96.55
tlinks	6122	95.38
TOTAL	**26635**	**96.53**

A TimeBank document can be seen as having three parts: the header, the text and the time and event descriptions (instances and links between temporal entities). The transfer procedure is designated for the header and the text parts only. The time and event descriptions make use of the `EVENT`, `TIMEX3` and `SIGNAL` IDs from the first two parts (see `MAKEINSTANCE`, `ALINK`, `TLINK` and `SLINK` tags). For these descriptions the transfer kept only those XML tags from the English version whose IDs belong to XML structures that have been transferred to Romanian. In Table 2 there is a statistic of the resulting Romanian TimeBank corpus in terms of all TimeML transferred markups.

5. Conclusions and Future Work

The research proves that the automatic import of the temporal annotations from English to Romanian is a worth doing enterprise with a very high success rate (in our experiments the transfer success rate was as high as 96.53%). The most important conclusion of the described work is that, as the manual annotation of the temporal expressions, events and their links is very time-consuming and expensive (Pustejovsky et al., 2002), the automatic transfer of annotations represents a solution, provided a parallel corpus involving the target language exists, the source language displays temporal annotation, and adequate processing tools are available.

In order to reach one of the initial goals – to have an English-Romanian parallel corpus, temporally annotated in both languages, as a basis for further researches – the annotation transfer has to be evaluated by using the manually corrected markups in the parallel corpus. This work permits to analyze the situations of perfect transfer and compare them with those situations in which:

- the temporal annotation transfer has to be done with some amendments when the temporal constructions in the two languages are not similar but they can be transferred using special developed rules, or
- it has to deal with language specific phenomena, such as the treatment of clitics or the PRO-drop phenomenon, specific to Romanian but not to English,
- or the transfer can not be performed (although, in our present corpus we have not identified such a situation).

This study opens the possibility to decide, based on corpus-evidence, how well the temporal theories can be applied to Romanian. In particular, the grammatical category of "aspect" of Romanian verbs could be better defined.

The temporal annotations can be used to create lists of Romanian triggers/non-triggers for specific purposes: categorization of verbs according to their class; noun-denoted and adjective-denoted events; specific temporal signals and expressions.

The best methods developed until now – machine learning-based or rule-based – will be studied, in order to create or adapt a temporal tagger – such as TARSQI (Verhagen et al., 2005) - for Romanian or even a language independent one.

The temporal annotated data together with time ontologies can be further used to represent the temporal structure of the discourse and its possible relations with other discourse structures, such as, for example, Rhetorical Structure (Mann and Thomson, 1987) or Veins Theory (Cristea et al., 1998; Forăscu et al., 2006).

The cooperation with specialists in the Natural Language Processing field will result in developing other specific applications, using various language and/or web resources: Reasoning with extracted temporal information, Temporal Summarization, Temporal Discourse Structure, Temporal Question-Answering, and Machine Translation.

6. References

[1] J.F. Allen. Towards a General Theory of Action and Time. Artificial Intelligence 23: 123-154. 1984.

[2] S. Armstrong. Multext: Multilingual Text Tools and Corpora. Lexikon und Text, pages 107–119, 1996.

[3] T. Brants. TnT – a statistical part-of-speech tagger. In Proceedings of the 6th Applied NLP Conference, ANLP-2000, Seattle, WA, pp. 224–231. 2000.

[4] B. Boguraev, R. Ando. Analysis of TimeBank as a Resource for TimeML Parsing. In Proceedings of LREC 2006, Genoa, Italy, pp. 71-76. 2006.

[5] A. Ceauşu. Integrated platform for Statistical Machine Translation system development (MTkit). Microsoft Imagine Cup. 2005.

[6] D. Cristea, N. Ide, L. Romary. Veins Theory. An Approach to Global Cohesion and Coherence. In Proceedings of COLING/ACL- 98, 281-285, Montreal, Canada. 1998.

[7] D. Cristea, C. Forăscu. Linguistic Resources and Technologies for Romanian Language. In Journal of Computer Science of Moldova, Academy of Science of Moldova, vol. 14, nr. 1(40), pp. 34-73. 2006.

[8] L. Ferro, L. Gerber, I. Mani, B. Sundheim, G. Wilson. TIDES 2005 Standard for the Annotation of Temporal Expressions, April 2005.

[9] C. Forăscu, I. Pistol, D. Cristea. Temporality in Relation with Discourse Structure. In Proceedings of LREC-2006, pp. 65-70, ISBN 2-9517408-2-4; Geneva, May 2006.

[10] J. Hobbs. Toward an Ontology for Time for the Semantic Web. In Proceedings of the LREC 2002 Workshop Annotation Standards for Temporal Information in Natural Language, p. 28-35, Las Palmas, Spain. 2002.

[11] J. Hobbs, J. Pustejovsky. Annotating and Reasoning about Time and Events, Proceedings of the AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning, Stanford, California. 2003.

[12] N. Ide, P. Bonhomme, L. Romary. XCES: An XML-based Encoding Standard for Linguistic Corpora. In Proceedings of the Second International Language Resources and Evaluation Conference., pp. 825-830. 2000.

[13] R. Ion.Word Sense Disambiguation Methods Applied to English and Romanian. (in Romanian). PhD thesis. Romanian Academy, Bucharest. 2007.

[14] G. Katz, F. Arosio. The Annotation of Temporal Information in Natural Language Sentences, in Proceedings of the ACL-2001 Workshop on Temporal and Spatial Information Processing, ACL-2001, Toulose, France, 6-11 July, p. 104-111. 2001.

[15] I. Mani, J. Pustejovsky, R. Gaizauskas (eds.). The Language of Time: A Reader. Oxford University Press, ISBN-13: 978-0-19-926853-5, May 2005.

[16] W. C. Mann, S. A. Thompson. Rhetorical structure theory: Description and con-struction of texts structures. In G. Kempen, editor, Natural Language Generation, 85–96. Martinus Nijhoff Publisher, Dordrecht. 1987.

[17] J. Martin, R. Mihalcea, T. Pedersen. Word Alignment for Languages with Scarce Resources. In Proceeding of the ACL2005 Workshop on "Building and Using Parallel Corpora: Datadriven Machine Translation and Beyond". Ann Arbor, Michigan, June, ACL, 65–74. 2005.

[18] J. Pustejovsky, L. Belanger, J. Castaño, R. Gaizauskas, P. Hanks, B. Ingria, G. Katz, D. Radev, A. Rumshisky, A. Sanfilippo, R. Sauri, A. Setzer, B. Sundheim, M. Verhagen. NRRC Summer Workshop on Temporal and Event Recognition for QA Systems, September 2002.

[19] J. Pustejovsky, M. Verhagen, R. Sauri, J. Littman, R. Gaizauskas, G. Katz, I. Mani, B. Knippen, A. Setzer. TimeBank 1.2. Linguistic Data Consortium. 2006.

[20] H. Reichenbach. The tenses of verbs, in H. Reichenbach (ed.), Elements of Symbolic Logic, Section 51, 287-298, New York: Macmillan. 1947.

[21] R. Sauri, J. Littman, B. Knippen, R. Gaizauskas, A. Setzer, J. Pustejovsky. TimeML Annotation Guidelines, Version 1.2.1, January 2006.

[22] A. Setzer. Temporal Information in Newswire Articles: an Annotation Scheme and Corpus Study, PhD dissertation, University of Sheffield. 2001.

[23] D. Tufiş, R. Ion, A. Ceauşu, D. Ştefănescu. Combined Aligners. In Proceedings of the ACL2005 Workshop on "Building and Using Parallel Corpora: Data-driven Machine Translation and Beyond", *Ann Arbor, Michigan, June,* Association for Computational Linguistics, pp. 107-110. 2005

[24] D. Tufiş, R. Ion, A. Ceauşu, D. Ştefănescu. Improved Lexical Alignment by Combining Multiple Reified Alignments. In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL2006), Trento, Italy, 3-7 April. pp. 153-160. 2006.

[25] D. Tufiş, A.M. Barbu. Revealing translators knowledge: statistical methods in constructing practical translation lexicons for language and speech processing. In International Journal of Speech Technology. Kluwer Academic Publishers, no.5. pp.199-20. 2002.

[26] M. Verhagen, I. Mani, R. Sauri, J. Littman, R. Knippen, S. Bae Jang, A. Rumshisky, J. Phillips, J. Pustejovsky. Automating Temporal Annotation with TARSQI. Short paper. In Proceedings of the 43rd Annual Meeting of the ACL, 81-84, Ann Arbor, USA. 2005.