Automatic Analysis of Temporal Relations within a Discourse

Petr Němec
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University
Malostranské náměstí 25, 11800 Praha, Czech Republic
nemec@ufal.mff.cuni.cz

Abstract

In this paper we present a rule–based system of automatic analysis of temporal relations within a discourse for Czech. The core algorithm of the system is the so–called "recursive temporal property" (RTP) of Czech that predicts relative ordering relations between finite verbs in a sentence based on the information provided only by the grammar. We correct and extend the RTP by various linguistically–oriented rules. Some of these rules represent a contribution to the existing grammar theory of temporal relations in Czech. The resulting system yields F–measure of 57.77–65.29.

1. Introduction

In this paper we present a rule–based system of automatic analysis of temporal relations within a discourse in Czech. Such a system can be used in variety of applications such as question–answering, information retrieval, text summarization, machine translation (where the knowledge of the proper sequence of events expressed in a sentence makes it possible to select the correct tense in target languages with rich tense system) and represents a step towards natural language understanding in general.

There has been a significant prior research on the temporality issues from various perspectives, e.g., by Hwang and Schubert [6], Kamp and Ryle [7], Lascarides and Asher [8], Hitzeman et al. [5], Panevová [17] for Czech, and many others.

Prior to experiments with automatic analysis, it is necessary to devise an annotation scheme and create a corpus annotated for temporality. The work on both the scheme and the corpus is described in [15] and is summarized in Section 2.

The core algorithm of the presented system is the so–called "recursive temporal property" (*RTP*) of Czech [17] that predicts relative ordering relations between finite verbs in a sentence based on the information provided by the grammar only. By the time RTP was formulated by Panevová there was no electronic (or temporal) corpus available, the possibilities to test RTP were therefore limited. Having a temporal corpus at our disposal, it is now possible to verify the RTP hypothesis to a greater extent.

As it turns out, there are actually many counterexamples to RTP. Some of them may be corrected and these revisions are represented by additional rules in our system. In some cases the correction touches the very basics of RTP and is of linguistic interest per se.

Moreover, RTP has been extended via other rules to draw additional relations. The system of automatic comparison of temporal expressions that anchors discourse events has been implemented which makes it possible to draw additional inferences.

The paper is structured as follows: Section 2 introduces the annotation scheme. Basic features of the annotated corpus are listed in Section 3. Section 4 provides brief description of the relevant points of Czech morphology and describes RTP in detail. The definition of evaluation metrics is subject of Section 5. The respective rules together with their performance are listed in Section 6. Section 7 compares the presented

approach to the existing related work and Section 8 concludes the paper.

2. Annotation Scheme Overview

2.1. Basic Principles

In accordance with Novák [14] we recognize starting time point anchor E_s and ending time point anchor E_e of each event (activity, accomplishment, state etc.) E expressed in a discourse. E_s anchors the beginning of the event whereas E_e anchors the time the event is finished. If an event E takes place in one single time point, we take $E_s = E_e$. These anchors are interpreted as time points on the real time axis.

The set of all the anchor pairs and the set of time–of–speech points (one for each discourse utterance) together form the *temporal space* of a discourse. Consider the following example:

1. *A consortium of private investors operating as BPH Funding Co. said yesterday that it could eventually make a $300 million cash bid.*

2. *Today it announced that it no longer considers the possibility.*

There are two time–of–speech points (*1.tos, 2.tos*) and starting and ending points for the events expressed by the words *operating* (*op.s, op.e*), *said* (*say.s, say.e*), *make* (*mk.s, mk.e*), *announced* (*anc.s, anc.e*), and *considers* (*cnsd.s, cnsd.e*).

The task of the temporal annotation of a discourse is to identify its temporal space and to determine relations between these points.

The following relative ordering relations may hold between two time points p and q: precedence ($p \prec q$), precedence–or–equality ($p \preceq q$), antecedence ($p \succ q$), antecedence–or–equality ($p \succeq q$) and equality ($p = q$). For the example mentioned above the basic possible set of relative ordering relations would be as follows:

- $1.tos \prec 2.tos$ (sentence order)

- $say.s = say.e \prec 1.tos$ (*said* expresses a single time point event)

- $1.tos \prec mk.s = mk.e$ (*make* takes place in the future if it takes place at all)

- $anc.s = anc.e \prec 2.tos$

- $cnsd.s \prec anc.s \prec cnsd.e$

These relations correspond to the information provided by the grammar. Nevertheless, more relations can be inferred with various level of confidence:

- $op.s \prec say.s$ (the consortium was probably operating as BPH Funding before it made a statement yesterday)

- $op.s \preceq cnsd.s$ (the same for considering the offer)

- $op.s \prec 1.tos \prec op.e$ (the consortium is probably operating as BPH Funding in the time the sentence was written)

- $say.s \prec anc.s$ (follows from the temporal expressions (*yesterday*, *today*))

- $op.s \prec 2.tos \prec op.e$ (the state of affairs is true even in the time–of–speech of the other sentence)

- $cnsd.s \prec 1.tos \prec cnsd.e$ (the same)

The complete annotation of relative ordering of the respective events is the transitive closure of the declared relations (e.g., from $a \prec b \prec c$ the $a \prec c$ relation is inferred).

The annotation of all the binary inter–sentential relations is very difficult as their number rises quadratically with the number of time points within the discourse. It is unfeasible for a human annotator to consider all of them. In practice, we annotate all the potential relations crossing the adjacent sentence boundary, i.e. when annotating a sentence we also consider all the time points of the previous one.

2.2. Time Expressions

Values of some time points are determined by functions of other time points or are even specified absolutely. For example, in the sentence

Last year we spent our holiday in Austria and it was very similar to our vacation in Germany in February 1980.

the event of spending the holiday in Austria is determined as a function (last year) of the time–of–speech point whereas the event of spending the holiday in Germany has been positioned absolutely to the interval of February 1980. Note that although we may not know when exactly the sentence was uttered or written we still understand it, should be able to annotate it and draw potential inferences from it.

To capture this kind of information we have developed a formalism based on the functions outlined below. It represents content of the expressions such as "last Friday", "beginning of the next month", "the middle of 80s" etc. Moreover, it allows for a construction of an efficient algorithm for the computation of partial ordering of these expressions on the real time axis.

We will only present the functions informally and in an abbreviated format:

- $entityRange(EntityType, Number)$ function returns the time range represented by $Number$ time entities of type $EntityType$, e.g., $entityRange(year, 2)$ returns time range of two years.

- $shift(Point, Distance)$ returns the time point that succeeds or precedes $Point$ by $Distance$, e.g., $shift(p, entityRange(year, -2))$ returns the time point corresponding to the expression "two years ago" relative to p.

- $span(Point, EntityType)$ function returns the concrete time interval of the time entity of type $EntityType$ that contains the time point $Point$, e.g., $span(p, month)$ returns the interval of the month containing time point p.

- The $find(Point, Entity, Index)$ function finds the $Index$–th occurrence of $Entity$ succeeding or preceding (if $Index$ is negative) $Point$. For example, $find(p, Monday, 1)$ returns the interval of the next Monday relative to p.

- The $part(Interval, Part)$ function returns the $Part$ of $Interval$, e.g., $part(1980|2, beginning)$ corresponds to the (vague) expression "beginning of the February 1980".

- The $series(Point, Entity, Count)$ function returns the set of intervals corresponding to the $Count$ successive entities $Entity$ relative to $Point$, e.g., $series(p, day, 2)$ corresponds to the expression "next two days" relative to p.

- Other auxiliary functions representing operations on intervals, arithmetic on time ranges, direct constructors of time points and time intervals etc.

These functions and operators may be arbitrarily composed to form the resulting expression. A time point can then be positioned inside (outside) the interval specified by this expression. E.g., the starting point $sp.s$ of $spend$ from the example can be positioned to "last year" as
$$sp.s \in$$
$$span(shift(tos, -entityRange(year, 1)), year)$$
or as
$$sp.s \in find(tos, year, -1)$$

where tos is the time of speech. The same mechanism can be used to specify absolute distance between two time points, etc.

2.3. Special Modifiers

Some discourse expressions do not express a single event taking place in a single interval of time. The event may be iterative as in *"Last month, I used to wake up every morning and run 10 miles."* and it is necessary to capture this information. However, it is not enough to mark each of the two events separately as the connection between them would be lost. Instead, we introduce the notion of a $plan$ to capture the fact that the events are recurrent and one follows after another in a "single run": the two events are annotated as usual except that they are declared to be part of one particular box – the plan. The plan itself has a starting time point and ending time point (plan boundaries) which denote the interval of plan's validity (in the example, these are directly specified as the starting and ending of the "last month"). Repeat period of the plan may also be specified ("one day" in our example).

Another special situation arises when the event occurs separately for each actor in distributive readings

such as *"Each company built its own headquarters in Boston."*. This is indicated in the annotation by a special marker. The two cases may co–occur as in *"Many people wake up every morning..."*. In this case the plan is marked as distributive as well.

3. Annotated Corpus

Although the annotation scheme itself is language independent and can be used to annotate plain texts, it is particulary convenient to link the temporal annotation with the existing level of tectogrammatic annotation within the framework of the Prague Dependency Treebank (PDT) [2].

The tectogrammatic representation (TR) of a sentence captures its deep–syntax properties and relations as a tectogrammatical tree structure (TGTS). A TGTS is a dependency tree, the nodes of which represent the autosemantic words of the sentence. Each node is labelled with an inflectionally reduced word–form called the lemma and a functor that describes the deep syntactic relationship to its governor (parent node) such as actor (ACT), patient (PAT), temporal specification (TWHEN), locative etc. Additionally, the nodes are labelled with grammatemes that capture further morphological and semantic information corresponding to the autosemantic words such as tense, aspect, gender, number, mood, modality etc. The ordering of the respective nodes captures the topic–focus articulation and the so-called scale of the communicative dynamism [4] which does not in general correspond to the surface ordering. An example of a simplified tectogrammatic tree is depicted in Figure 1. For a detailed description of the TGTS annotation scheme see [2]. The temporal annotation can be therefore viewed as a natural extension to the PDT framework.

We have annotated the Czech translation of portion (parts of sections 22, 23 and 24) of Wall Street Journal (WSJ) as present in the parallel Prague Czech–English Dependency Treebank (PCEDT) corpus [21][1]. Under the PCEDT project these sections have been translated

---

[1]The primary reason for the decision to use these data for temporal annotation (in contrast to annotating the original PDT data) was the possibility to integrate the resulting automatic temporal information retrieval system into this machine translation framework. This allows us to directly evaluate the contribution to the existing machine translation system.
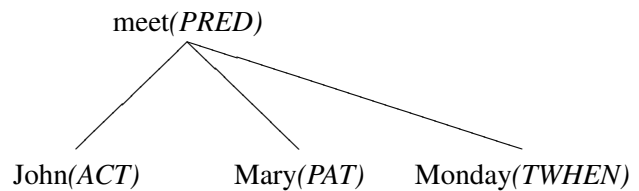

Figure 1. A simplified tectogrammatic tree representing the sentence "John met Mary on Monday." Only the functors are displayed.

and annotated at the tectogrammatical level for purposes of machine translation [3]. Currently, the development testing data set (233 sentences) and the evaluation testing data set (231 sentences) are annotated for temporality and used as testing data for our experiments. Detailed statistics on the distribution of the respective annotated entities (verbs, adjectives and nouns) can be found in [15].

As there were many errors in the relevant grammatemes of the tectogrammatical annotation (tense, aspect) within PCEDT, these have been manually corrected.

All the experiments described in this paper assume that TGTSs for the respective sentences are available. This assumption makes it impossible to analyze plain texts directly, they have to be parsed and their TGTS has to be built first.

4. Recursive Temporal Property

To ease the understanding of the subsequent text, we give a very brief overview of the relevant points regarding morphologic system of Czech (which is substantially different from, for example, English tense morphology).

In Czech, a verb bears either perfective or imperfective aspect. The former corresponds primarily to a complex event (taking place in a single time point), the latter corresponds primarily to a procesual event (event or state with a time duration). The aspect information is therefore encoded lexically (in contrast to English where it is encoded gramatically). The following example demonstrates the difference:

Petr dělal domácí úkol.   (Petr was doing his

COMPUTER SOCIETY

homework.)
Petr udělal domácí úkol. (Petr has done his home-work.)

A finite verb bears one of the three basic tenses – past, present or future. There are no perfect tenses. A perfective verb can only bear past or future tense.

Another important notion is the so–called "content clause". Roughly speaking, content clauses correspond to indirect speech and some similar types of clauses. Object clauses form the core of content clauses. See [16] for a detailed analysis of this notion.

RTP may be described as follows: The morphological tense of a finite verb (A) – past, present, future – determines its temporal relation – precedence, overlap (i.e. at least one common point), antecedence – with its reference point.

- The reference point of the matrix clause of the sentence is the time–of–speech of the sentence.

- The reference point of the content clause is the point of the event of its governing clause.

- The reference point (R) of the other (subordinated) clause is determined as identical with the reference point of its governing clause.

For example, consider the sentence

"Honza řekl, že Petr našel řešení, které Marie schvaluje."
(lit. Honza said that Petr found solution that Marie approves.)

whose schematized dependency tree is depicted in Figure 2 (M denotes the matrix verb, O is the head of a content (object) clause and A is the head of an attributive clause. C and P denote a complex and a procesual verb, respectively.) According to RTP

- *řekl* precedes time–of–speech, thus
  řekl.s (=řekl.e) $\leq$ *tos*

- *naš*el precedes *řekl*, thus
  našel.s (=našel.e) $\leq$ *řekl.s (=řekl.e)*

- *schvaluje* is simultaneous (overlaps) with *řekl*, thus
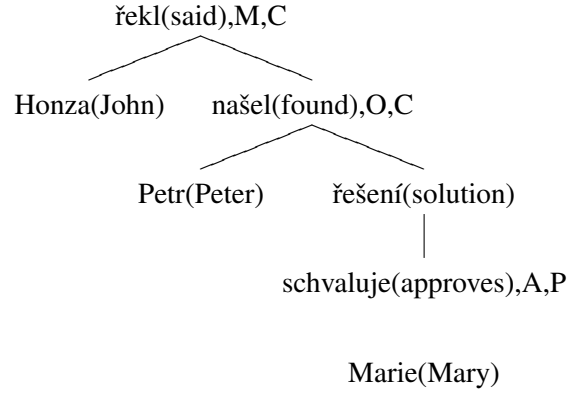 *schvaluje.s* $\leq$ *řekl.s (=řekl.e)* $\leq$ *schvaluje.e*

Figure 2. A schematized dependency tree representing the sentence "Honza řekl, že Petr našel řešení, které Marie schvaluje.". (Surface order is not preserved and auxiliary words are missing.)

5. Evaluation Metrics

The evaluation of the performance of the presented system is based on the comparison between the reference (annotated test set) matrix and the hypothesis matrix of temporal relations for each discourse (WSJ article) respectively. The transitive closure of both is computed before the evaluation takes place.

We compute precision P and recall R defined as

$$P = \frac{\#Correct}{\#Hypothesis}$$
$$R = \frac{\#Correct}{\#Reference}$$

where $\#Correct$ denotes the number of correctly determined relations, $\#Hypothesis$ is the number of determined relations and $\#Reference$ is the number of relations within annotation.

A relation between two points (p and q) is considered correct if its value (i.e. one of $\prec, \preceq, \succ, \succeq, =$) is exactly the same within the hypothesis and the reference. If the hypothesis contains a weak relation ($\preceq, \succeq$) instead of the (correct) corresponding "strong" relation ($\prec, \succ$, or $=$) we call the result "weakly correct". We denote P_w and R_w the versions of P and R that consider weakly correct relations as correct. By default, if p or q is marked as distributive and unrecognized

as such or is part of an unrecognized plan the relation is considered incorrect. We also denote P_f and R_f the versions of P and R that do not take these markers into account[2]. Finally, P_{wf} and R_{wf} combine both aspects and are therefore the most permissive metrics. Due to space limitations we provide only P, R, P_{wf} and R_{wf} (and the corresponding F–measure rates F and F_{wf}) in the evaluation tables.

6. Automatic Analysis

In this section we list the respective rules used in the system of automatic analysis of temporal relations together with their performance. We also list counterexamples to RTP we were unable repair due to their complexity.

Table 1 gives complete overview of the system performance on the development testing set: the unrefined RTP baseline, the performance with all rules on and the contribution of each respective rule by switching it off and providing difference to the all–on state. (We do not provide difference of each single rule against the RTP baseline because the rules are not independent – application of some of them requires prior application of others to yield correct results.) The abbreviations used for the respective rules in Table 1 are introduced in the following subsections together with the description of the rule.

The errors that remain after the application of all the described rules are caused partly by counterexamples to RTP we were unable to correct (as described in subsections 6.3 and 6.5), partly by (common) errors in the structure (or grammatemes) of the underlying TGTS and partly by the errors in the temporal annotation.

Table 2 list the overall performance of the system on the evaluation testing set. As we were unable to process deverbative adjectives and nouns (as explained in the subsection 6.8) we also report the performance measured only on verbs in Table 3.

---

[2]We provide these supplementary metrics as well because we believe that providing a weak (defensive) analysis or a correct analysis on an unrecognized plan or distributive marker is still an achievement that should be taken into account. On the other hand, it is obvious that in these cases the system did not succeed entirely.

All	P	R	F	P_{wf}	R_{wf}	F_{wf}
	75.29	41.64	**53.62**	84.7	46.84	**60.32**
-Rule	dP	dR	dF	dP_{wf}	dR_{wf}	dF_{wf}
HP	-3.19	-0.99	**-1.63**	-3.79	-1.22	**-1.98**
CCA	-0.33	-1.63	**-1.45**	-0.35	-1.82	**-1.61**
ZC	-2.46	-0.51	**-1.05**	-2.56	-0.45	**-1.03**
DS	-1.59	-0.13	**-0.52**	-1.57	-0.02	**-0.42**
INF	0.61	-0.47	**-0.24**	0.57	-0.59	**-0.35**
NCV	-0.16	-0.21	**-0.21**	-0.13	-0.21	**-0.21**
TE	-0.07	-0.22	**-0.20**	-0.04	-0.22	**-0.19**
CC	-0.49	-0.08	**-0.19**	-0.39	0.00	**-0.10**
TC	-0.12	-0.08	**-0.10**	0.00	-0.02	**-0.02**
RTP	-5.78	-4.19	**-4.95**	-6.34	-4.62	**-5.45**

Table 1. The overall performance of the system with all rules on the development testing set and the list of differences to this performance for each respective rule off. The last row corresponds to the unrefined RTP algorithm (baseline).

	P	R	F	P_{wf}	R_{wf}	F_{wf}
RTP	70.77	44.17	**54.39**	79.24	49.46	**60.90**
Rules	72.32	48.09	**57.77**	81.74	54.35	**65.29**

Table 2. The overall performance of the system with all rules on and off (RTP baseline), respectively, on the evaluation testing set.

	P	R	F	P_{wf}	R_{wf}	F_{wf}
RTP	70.77	61.32	**65.71**	79.24	68.66	**73.57**
Rules	72.32	66.75	**69.42**	81.74	75.44	**78.46**

Table 3. The overall performance of the system with all rules on and off (RTP baseline), respectively, on the evaluation testing set taking only the verbs into account.

6.1. Direct Speech

If a sentence is part of a direct speech segment its reference point (i.e. the reference point of its matrix verb) obviously is not the time–of–speech point (as assumed by the unrefined RTP) but rather the speech verb introducing the direct speech segment. A TGTS – one per sentence – provides by itself no clue as to whether the sentence it represents belongs to a direct speech segment or not. To correct the produced errors it is necessary to detect direct speech segments (based on quotation markers) and speech verbs. The rule is denoted as DS (Direct Speech) in Table 1.

6.2. Content Clauses

It turns out that – in contrast to the assumptions of the RTP – certain objective clauses show "non–content" behavior in the context of temporal relations and vice versa.

The objective clauses conjoined with the governing clause by the "co" ("what") pronoun behave like non–content clauses:

"Honza udělá, co bude považovat za nejlepší."
(lit. Honza will_do what (he) will consider for best.)

In this example, *bude považovat (will consider)* does not follow after *udělá (will_do)* as predicted by the RTP but is rather simultaneous (overlaps) with it. However, one might argue that the clause is rather an attributive one as it is an ellipsis of "Honza udělá to, co..." (lit. Honza will do that, what ...).

The inverse situation is represented by the adverbial subordinated clauses conjoined by the conjunction "aby" (so that):

"Pan Shidler řekl, že firma propustila zaměstnance, aby ušetřila."
(lit. Mr. Shidler said that company fired employees so that (it) saves (money).)

According to the RTP, the reference point of the event *ušetřila (saves[3])* is the point of *řekl (said)*, but there is no relation whatsoever – the company may

---

[3]This is actually a conditional in Czech so its tense meaning is not clear.

---

save the money before or after Mr. Shidler made the statement. Instead, the event in the subordinate clause follows the event – *propustila (fired)* – in the governing clause, which is the behavior of a content clause. It seems that it is the general behavior of the aim clauses.

The rule that reflect the proper behavior of both types of clauses is denoted as CC (Content Clauses) in Table 1.

6.3. Aspect Deviations

Primarily, the verbal aspect corresponds to the duration of the denoted event: complex verbs denote events taking place in one single time point (with certain level of approximation) while procesual verbs denote events that last over a certain time period. Occasionally, this correspondence is broken. We have identified at least three cases where it happens systematically.

The first case is the so–called "historic present" (already described in this temporal context by Panevová in [17]) as in

"Pan Shidler v rozhovoru říká, že společnost změní svou investiční strategii."
(lit. Mr. Shidler in interview says that company will_change its investment strategy.)

where *říká (says)* actually means *řekl (said)* and the procesual present is used for stylistic reasons. Note, however, that sentences such as

"Zpráva říká, že Martin přijede zítra."
(lit. Message says that Martin will_come tomorrow.)

are not instances of the historic present. We have attempted a correction of errors that stem from this phenomenon for speech verbs only. To do so, it is necessary a) to detect an animate actor of the speech verb, b) to ensure that the instance really is a single point event and not a recurrent event as in

"Maminka říká, že nemám chodit do lesa."
(lit. Mama says that I should not go to forest).

As far as a) is concerned, we use a simple animate actor detection algorithm: if the actor is not

a proper noun it consults the Czech EuroWordNet ontology [22] whether the actor is an animate entity. If the actor is a proper noun, the algorithm scans for all the occurrences of that name within the discourse and checks whether there is an associated description (via apposition, i.e. *"Tom Baker, president of Machinists District 751"*) that is again checked against the ontology. We have not found any examples of the mentioned recurrent event usage in our domain of newspaper articles (WSJ) so we have not even attempted to fulfil the second condition. The current solution may be therefore viewed as a domain–specific rule (denoted as HP (Historic Present) in Table 1).

The second case where the verb's aspect may not correspond to the nature of the event is the negation of a complex verb:

"Boeing zásilku zatím nedostal."
(lit. Boeing package(accusative) so far not got.)

obdržet (to receive) is a complex verb but its negation in this context denotes a progressive state. However, in context such as *"Boeing zásilku včera v 16:00 nedostal." (lit. Boeing package(accusative) yesterday in 16:00 not got.)* the event is still taking place in a single point (16:00 yesterday). Currently, we detect the presence of the specific adverbs such as *zatím* or *dosud (so far)* to correct this type of errors. The rule is denoted as NCV (Negated Complex Verb) in Table 1.

Another (common) situation that leads to aspect deviations is the appurtenance of an event to a plan. The procesual aspect of a verb can (but does not have to) be used to express iteration of a complex event. If this is the case an error occurs. Iterative events cause many errors in general as the nature of the described recurrent series of events is different.

Currently, we indicate the appurtenance of an event to a plan by detecting relevant temporal adverbs such as *obvykle (usually)*, *často (often)*, etc., but this handles only a fraction of all plan occurrences so there is essentially no improvement.

6.4. Conditionals

Although the conditional in Czech has two sets of forms (past conditional and present conditional),

this distinction is not present in the annotation of the PCEDT corpus. The information on the tense of a conditional expression – required by the RTP – is therefore not available. Worse yet, as the autosemantic part of the conditional expression is identical to the ordinary past tense, the tense attribute of the conditionals in the PCEDT corpus is incorrectly set to past. This naturally produces many errors.

The decision on the proper "content" tense of the event denoted by a conditional seems to be hard. Consider the following examples:

- (future) *Ty projekty jsou velké, ale nešli bychom do nich sami.*

  (lit. Those projects are big but (we) would_not_go into them alone.)

- (present) *Nezdá se, že by stávka měla nějaký efekt.*

  (lit. It_does_not_seem that strike would have any effect.)

- (past) *Přišel by, ale ona nechtěla.*

  (lit. He_would_come but she did_not_want (him to).)

We were unable to find any reliable principle that would allow us to distinguish between the respective cases. We therefore do not determine any relation when considering a conditional – the rule is denoted as ZC (Zero Conditionals) in Table 1.

It is possible to set, for instance, future defaults and gain a higher F–measure as the precision loss is less than the recall gain but we have decided not to as the rules should be as reliable as possible.

6.5. Other Exceptional Clauses

We have identified two more interesting sources of errors. The first one is represented by sentences containing certain clauses where two different tense forms could be used: the future and the present tense. Consider the following example:

"Otec chce synovi předat (namísto současné trosky) podnik, který prosperuje (= bude prosperovat)."
(lit. Father wants (on his) son pass (instead_of current

ruin) company that prospers (= will prosper).

Future tense corresponds to the "real" meaning of the clause predicted by the RTP, whereas the present tense is a marked variant yielding incorrect results in the context of the RTP.

The other case (described already by Panevová in [17]) is the so–called "commentary" of the speaker as in

"Pan Shidler řekl, že společnost chce odkoupit pozemky v Ohiu, které nyní vláda plánuje rozdělit."
(lit. Mr. Shidler said that company wants (to) buy properties in Ohio that now goverment plans (to) divide.)

where the last subordinate clause is not part of the reported speech but rather a commentary of the author of the article.

We do not see any feasible way to identify and correct these instances.

6.6. Common Complex Anchor

If the reference point of a present procesual verb (A) is again a procesual verb (B) it is not possible to determine any relations in terms of the starting and the ending points of the two events because they may overlap in a variety of ways. If B is in the present tense we have tried to extend RTP by taking the reference point of B (C) as the reference point of A. If C is again a procesual present verb we use its reference point (D) and so forth. For example, in

"Pavel řekl, že si myslí, že pracuje dobře."
(lit. Pavel said that he thinks that he works well.)

this principle allows us to determine the overlap relation between *pracuje (works)* and *řekl (said).*

We have encountered no errors due to this extension. A potential counterexample would consist in the possible intransitivity of the overlap relation.

This rule is denoted as CCA (Common Complex Anchor) in Table 1.

6.7. Infinitives

We have discussed the relations between finite verbs (and infinitives surrounded by auxiliary finite verbs) so far. Let us consider the situation where an infinitive (I) is dependent on an autosemantic finite verb (F). Our hypothesis is as follows: If I is complex then $F.s \leq I.s$. (e.g. *"Přišel nám pomoci." (He came to help us.)*).

We are aware that this is quite an ad-hoc hypothesis, nevertheless we have not found any real counterexample to it in the corpus. Note that the hypothesis obviously does not hold if I is procesual as the following examples demonstrate:

"Odmítl nadále pracovat."
(He denied to work further.)

"To dělníkům pomáhá vyvíjet tlak na společnost."
(That helps workers to put pressure on the company.)

This rule is denoted as INF (Infinitives) in Table 1. Note that this rule actually decreases precision but this is due to errors caused indirectly (via transitive closure) by some other fundamental error. We have therefore decided to keep this rule.

6.8. Deverbative Adjectives and Nouns

In Czech, there are two basic categories of deverbative adjectives: "active' (corresponding to the English -ing verb form) and "resultative" (corresponding to the English perfect verb form). Both seem to be counterparts of the subordinated attributive clauses. Unfortunately, a reliable determination of a temporal relation for any of these adjectives does not seem to be possible as the following examples demonstrate:

- *Řekl, že (namísto současné trosky) předá synovi prosperující podnik. (= podnik, který bude prosperovat)*

  (He said that (instead of the current ruin) he would pass a well-performing company on his son.) (= a company that will perform well)

- *Slíbil jí, že její prosperující podnik neprodá. (= podnik, který prosperuje)*

(He promised her that he would not sell her well-performing company.) (= a company that performs well)

- *Policie prohlásila, že ukradené zlato bude navráceno bance. (= zlato, které bylo ukradeno)*

  (The policie announced that the stolen gold will be returned to the bank.) (= gold that was stolen)

- *Před loupeží slíbil, že ukradené peníze budou rozděleny rovným dílem. (= peníze, které budou ukradeny)*

  (He promised before the robbery that the stolen money would be distributed evenly.) (= money that will be stolen)

6.9. Temporal Clauses

Czech temporal subordinate clauses determine the relative ordering relation of the expressed event not only to its reference point (according to the RTP) but also to its governing clause. TGTS captures various types of temporal complements such as "before", "after", "parallel to" etc. For certain types of complements, this information can be quite straightforwardly used to determine relations between the head verb of a temporal subordinate clause and the head verb of its governing clause.

The appropriate rule is denoted as TC (Temporal Clauses) in Table 1. The F–measure increase is very low due to small number of affected temporal clauses.

6.10. Inference from Time Expressions

As shown in Section 2.2, the annotation scheme provides a functional approach to capture meaning of various time expressions. In order to infer relative ordering relations based on these determinations it is necessary to construct the functional composition and to compare the respective compositions.

The construction of the compositions is handled by a parser module that scans TGTSs for occurrences of subtrees that are accepted by the given tree grammar. The functional composition is then built incrementally. The majority of simple expressions such as "two days ago", "tomorrow", "in the beginning of the next week", "in 1987", "in the end of the last summer"

etc. are accepted. The description of the grammar together with the list of accepted expressions is beyond scope of this paper.

Two compositions are compared against each other by another module that traces the composition inside–out and uses a "beam" structure to keep record of the distance (or absolute value) of the outcome of the last visited function to the source interval or point. If the source point is a variable (the module is currently limited to compositions with at most one variable) only the estimates are provided while tracing the composition. If the compositions are comparable (e.g., the highest estimate carried by the beam of one composition is lower or equal to the lowest estimate carried by the beam of the other composition) then the ordering of the events anchored by the corresponding time expression may be inferred.

Use of this mechanism is denoted as TE (Time Expressions) in Table 1. The performance increase is low due to the fact that only the inferences based on time expressions within one sentence or in adjacent sentences may be drawn as the annotation does not contain relations between events across more than one sentence – we would not be able to evaluate full analysis based on time expressions. There are only a few adjacent time expressions in the corpus our system is able to compare. This is unfortunate as it prevents us from making use of much of the available information (at least for evaluation purposes).

7. Related Work

There has been significant prior research on the extraction of temporal information in different languages.

Examples of rule–based approaches are that by Mani and Wilson [13] for English, Schilder and Habel [20] for German and Li et al. [10] for Chinese. They report varying accuracy rates (59.4%, 84%, 93% respectively).

Machine–learning approaches to the problem emerge, too. Mani et al. [11] trained a decision tree for the problem of anchoring events to reference times yielding 80.2 F–measure. Later, Mani et al. [12] used a Maximum Entropy classifier on the union of the Time-Bank corpus [18] and the Opinion Corpus. The features they used were similar to those we were dis-

IEEE
COMPUTER
SOCIETY

cussing in this paper (such as aspect, tense, negation, modality etc.). The classifier yields results as high as 93% accuracy which is a significant improvement over the baseline represented by a rule–based approach. There is also work of Boguraev and Ando [1] who trained a classifier on Timebank corpus for event anchoring for events and times that yielded 53.1 F–measure. Li et al. [9] obtained approximately 80% accuracy on ordering temporal relations in Chinese texts.

Because these success rates are measured on different testing data, use different annotation schemata (such as TimeML [19] used by the cited work of Mani et al. that treats events quite differently than our annotation scheme) and evaluation metrics, it is difficult to provide a direct comparison of the achieved results. Nevertheless, it seems clear that machine–learning approaches can highly increase the performance of a system of automatic temporal analysis. Mani et al. [12] is a fine example of this. Many of the problems mentioned in Section 6 (conditionals, deverbative adjectives etc.) as well as many unresolved relations could be possibly resolved by such an approach.

8. Conclusion

We have presented a rule–based system of automatic analysis of temporal relations within a discourse together with the description of the respective rules. The system yields 72.32%–81.74% precision and 48.09%–54.35% recall which corresponds to 57.77–65.29 F–measure.

We have focused our attention mainly on the creation of reliable rules with no apparent counterexamples. We have shown that these rules have in general little impact on the overall performance of the system, although some of them are of linguistic relevance as they provide correction to the existing linguistic theory.

Although the achieved results are not directly comparable to those of the other published methods, we believe that some of these approaches – especially those based on machine–learning – provide superior performance in comparison to our system. As mentioned in the previous section, there are several areas whose processing by a machine learning method could bring significant improvement.

9. Acknowledgements

References

[1] B. Boguraev and R. K. Ando. Timeml-compliant text analysis for temporal reasoning. *Proceedings of IJ-CAI*, pages 997–1003, 2005.

[2] A. Bohmová, J. Hajič, E. Hajičová, and B. Vidová-Hladká. The prague dependency treebank: Three-level annotation scenario. *Treebanks: Building and Using Syntactically Annotated Corpora*, 2002.

[3] J. Hajič, M. Čmejrek, B. Dorr, Y. Ding, J. Eisner, D. Gildea, T. Koo, K. Parton, D. Radev, and O. Rambow. *Natural language generation in the context of machine translation*. Technical report, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, 2002.

[4] E. Hajičová and P. Sgall. The ordering principle. *Journal of Pragmatics*, pages 435–454, 1987.

[5] J. Hitzeman, M. Moens, and C. Grover. Algorithms for analyzing the temporal structure of discourse. *Proceedings of EACL*, pages 253–260, 1995.

[6] C. H. Hwang and L. K. Schubert. Tense trees as the fine structure of discourse. *Proceedings of ACL*, pages 232–240, 1992.

[7] H. Kamp and U. Ryle. *From Discourse to Logic (Part 2)*. Kluwer Academic Publishers, Dordrecht, 1993.

[8] A. Lascarides and N. Asher. Temporal relations, discourse structure, and commonsense entailment. *Linguistics and Philosophy*, (16):437–494, 1993.

[9] W. Li, K.-F. Wong, G. Cao, and C. Yuan. Applying machine learning to chinese temporal relation resolution. *Proceedings of ACL*, pages 582–588, 2004.

IEEE
COMPUTER
SOCIETY

[10] W. Li, K.-F. Wong, and C. Yuan. Toward automatic chinese temporal information extraction. *JASIST*, (52):748–762, 2001.

[11] I. Mani, B. Schiffman, and J. Zhang. Inferring temporal ordering of events in news. *Proceedings of HLT-NAACL*, pages 55–57, 2003.

[12] I. Mani, M. Verhagen, B. Wellner, C. M. Lee, and J. Pustejovsky. Machine learning of temporal relations. *Proceedings of ACL*, pages 753–760, 2006.

[13] I. Mani and G. Wilson. Robust temporal processing of news. *Proceedings of ACL*, pages 69–76, 2000.

[14] V. Novák. Towards logical representation of language structure. *The Prague Bulletin of Mathematical Linguistics*, (82):5–86, 2004.

[15] P. Němec. Annotation of temporal relations within a discourse. *Proceedings of Text, Speech and Dialogue*, pages 181–188, 2006.

[16] J. Panevová. Vedlejší věty obsahové. *Slovo a slovesnost*, (4), 1971.

[17] J. Panevová, E. Benešová, and P. Sgall. *Čas a modalita v češtině.* Charles University, Prague, 1971.

[18] J. Pustejovsky, P. Hanks, R. Sauri, A. See, D. Day, L. Ferro, R. Gaizauskas, M. Lazo, A. Setzer, and B. Sundheim. The timebank corpus. *Corpus Linguistics*, pages 647–656, 2003.

[19] J. Pustejovsky, B. Ingria, R. Sauri, J. Castano, J. Littman, G. Rob, A. Setzer, G. Katz, and I. Mani. The specification language timeml. *I.Mani, J. Pustejovsky, and R. Gaizauskas, (eds.), The Language of Time: A Reader*, 2005.

[20] F. Schilder and C. Habel. From temporal expressions to temporal information: semantic tagging of news messages. *I.Mani, J. Pustejovsky, and R. Gaizauskas, (eds.), The Language of Time: A Reader*, 2005.

[21] M. Čmejrek, J. Cuřín, and J. Havelka. Prague czech-english dependency treebank: Any hopes for a common annotation scheme ? *Proceedings of HLT-NAACL Workshop: Frontiers in Corpus Annotation*, pages 47–54, 2004.

[22] P. Vossen. *EuroWordNet: a multilingual database with lexical semantic networks for European Languages.* Kluwer Academic Publishers, Dordrecht, 1999.