Analyzing Time-Related Clauses in Transparent Intensional Logic

Aleš Horák, Miloš Jakubíček, Vojtěch Kovář

Faculty of Informatics
Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic
{hales, jak, xkovar3}@fi.muni.cz

Abstract. The Normal Translation Algorithm (NTA) for Transparent Intensional Logic (TIL) describes the key parts of standard translation from natural language sentences to logical formulae. NTA was specified for the Czech language as a representative of a free-word-order language with rich morphological system.
In this paper, we describe the implementation of the sentence building part of NTA within a module of logical analysis in the Czech language syntactic parser synt. We show the respective lexicons and data structures in clause processing with numerous examples concentrating on the temporal aspects of clauses.

Key words: Transparent Intensional Logic, TIL, Normal Translation Algorithm, NTA, logical analysis, temporal analysis, time-related clauses

1 Introduction

Logical representation of a natural language sentence forms a firm basis for semantic analysis, machine learning techniques and automated reasoning. In current practical systems all these fields build on propositional and first-order logic due to the advantages and simplicity of their processing [1]. However, it can be shown that first-order formalisms are not able to handle systematically the natural language phenomena like intensionality, belief attitudes, grammatical tenses and modalities (modal verbs and modal particles in natural language).

In the following text, we describe a system theoretically based on the formalism of the Transparent Intensional Logic (TIL [2]), an expressive logical system introduced by Pavel Tichý [3], which works with a complex hierarchy of types, system of possible worlds and times and an inductive system for derivation of new facts from a knowledge base in development.

The actual implementation of the system works together with the system for syntactic analysis of the Czech language named *synt*. The parsing mechanism in `synt` [4,5] is based on the meta-grammar formalism working with a grammar of about 250 meta-rules with contextual actions for various phrase and sentence level tests. The Normal Translation Algorithm (NTA [6]) as a standard way of logical analysis of natural language by means of the compositions of syntactic

Aleš Horák, Pavel Rychlý (Eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2011, pp. 3–9, 2011. © Tribun EU 2011

constituents such as verb phrase, noun phrase, clause and their modifiers was provided by a basic prototype implementation. Current works on NTA have supplemented the process with high-coverage lexicons and new logical tests and rules with growing coverage on common news paper texts.

2 Logical Analysis on the Clause Level

The central point of each clause (as a "simple" sentence) is formed by the verb phrase, which represents the predicative skeleton of the clause. Tichý comes with a classification of significant verbs into two groups according to the classification of their meaning:

1. *attributive verbs* express what *qualities* the attributed objects *have* or what the objects *are*. Typical examples are sentences with adjectival predicates, such as

   Zmíněné zařízení je číslicově řízené.
   (The mentioned device is numerically controlled.)
   $\lambda w \lambda t (\exists i) \Big([[\bar{\textbf{číslicově}}_{wt}, \bar{\textbf{řízený}}], i] \ \wedge \ [[\textbf{zmíněný}, \textbf{zařízení}]_{wt}, i] \Big) \ldots \pi$

   The appropriate analysis of the attributive verbs lies in a proposition that ascribes the alluded property to the subject.
2. *episodic verbs*, on the other hand, express *actions*, they tell what the subject *does*:

   Celková úmrtnost klesá.
   (The overall mortality rate decreases.)
   $\lambda w_1 \lambda t_2 (\exists x_3)(\exists i_4) \Big([\textbf{Does}_{w_1 t_2}, i_4, [\textbf{Imp}_{w_1}, x_3]] \ \wedge \ x_3 = \textbf{klesat}_{w_1} \ \wedge$
   $[[\textbf{celkový}, \textbf{úmrtnost}]_{w_1 t_2}, i_4] \Big) \ldots \pi$

The main difference between attributive and episodic verbs consists in their time consumption — the attributive verbs do not take the time dimension into account, they just describe the *state* of the subject in the very one moment by saying that the subject has (or has not) a certain property. The analysis of episodic verbs is defined with the means called *events* and *episodes*, which are in detailed described in [7,8,6]. An example of the resulting analysis uses three functions — **Does** and **Imp**/**Perf** which describe the relation between the subject and the predicate in the sentence and display the verb's aspect.

In natural language, the time constructions are encoded with two principle means — the verb tense and time adverbial groups that serve as modifiers of the default verb meaning.

The contemporary Czech language has three verb tenses — the present tense, the past tense and the future tense. For the basic verb tense analysis, we can adapt the definitions specified for English in [7]. Unlike English the Czech verbs have the capability of expressing the verb aspect built directly into their grammatical category — every verb is exactly in either the imperfective or the perfective form. A special property that goes along with the perfective aspect is that these verbs do not form the present tense — they have only the past tense and the future tense forms.[1] The present tense of the perfective verbs is usually expressed by their imperfective counterparts.

The verb tenses that allow us to assert propositions with respect to the past or to the future can be looked at as mirror images of each other. In the analysis, we understand the past tense as a certain operation working over

1. the *underlying proposition* in the present tense form and
2. the *reference time span*
3. with regard to an *assertion moment*

The meaning of a past tense proposition is often connected not only with a certain reference time span (in the indefinite case the object Anytime) but also with a *frequency adverb* specifying how many times the proposition happened to be true. A frequency adverb is analysed as a $((o(o\tau))\pi)_\omega$-object, i.e. as a world-dependent operation that takes a proposition p to the class of time intervals that have the requested qualities regarding the chronology of p. For instance, the adverb 'dvakrát' (twice) takes every proposition p to a class of time intervals that have exactly two distinct intersections with the chronology of p. If the frequency adverb is not specified in the sentence, we assume that the frequency of the proposition is 'at least once' (object Onc) in its time span, which means that we do not limit the sentence's time span in that case.

Thus, if P denotes the past tense function a general schema of the logical analysis of a past tense sentence looks like

$$\mathsf{P}\Big(<\text{frequency adverb}>(<\text{proposition}>),\ <\text{reference time span}>\Big)$$

The simple past is best seen as a time-dependent relation between $(o(o\tau))$-objects and $(o\tau)$-objects which holds for those cases where the (past part of the) reference time span belongs to the acceptable classes of time moments as obtained by the frequency modification of the proposition's chronology. An example analysis of a sentence in the past tense is

Celní správa vyžadovala originální certifikát.
(Customs administration required an original certificate.)

$$\lambda w_1 \lambda t_2 \Bigg[\mathbf{P}_{t_2}, \Bigg[\mathbf{Onc}_{w_1}, \lambda w_3 \lambda t_4 (\exists x_5)(\exists i_6)(\exists i_7)\Big(\big[\mathbf{Does}_{w_3 t_4}, i_7, [\mathbf{Imp}_{w_3}, x_5]\big] \wedge$$
$$\wedge \big[[\mathbf{originální}, \mathbf{certifikát}]_{w_3 t_4}, i_6\big] \wedge x_5 = [\mathbf{vyžadovat}, i_6]_{w_3} \wedge$$
$$\wedge \big[[\mathbf{celní}, \mathbf{správa}]_{w_3 t_4}, i_7\big]\Big)\Bigg], \mathbf{Anytime}\Bigg] \ldots \pi$$

The future tense may be now defined analogously to the simple past tense as $\mathsf{F}/(o(o(o\tau))(o\tau))\tau$.

[1] e.g. 'začal/začne' (he started/he will start), 'zabil/zabije' (he killed/he will kill) or 'udělal/udělá' (he did/he will do).

3 Clauses Combined in Complex Sentences

The logical analysis described in the previous section did not exceed the boundaries of a single clause. Now, we are going to show how we can combine the subconstructions corresponding to particular clauses in order to capture the appropriate subordinate and coordinate relations between the clauses in the resulting construction of the whole sentence.

An important feature of a connected (subordinate or coordinate) clause is the way (usually the conjunction) how this clause connects to the principal clause. We use the clause conjunction as a leading element for determining the kind of the clause (and thus the form of its analysis).

For each syntactic rule which connects a clause with a (possible complex) sentence, the clauses conjunctions are extracted and a lexicon of *sentence schemata* is searched for the corresponding logical analysis. A thorough list of possible conjunctions and the corresponding logical analyses can be found in [6, pp.106–113], but here we will rather demonstrate the whole process on the actual implementation of this part of NTA. Let us take an example sentence

Petr včera přišel, když Markéta telefonovala babičce.
(Petr came yesterday, when Markéta phoned her grandmother.)

The conjunction 'když' (when/if) can have two readings in a sentence: a) the temporal *when*, or b) the implication *if* (*if then* is translated as 'když pak'). The logical analysis offers ambiguously both readings. The analysis process first find the corresponding constructions for both the simple clauses:

Petr včera přišel (Petr came yesterday):
`time span TIL:` $\lambda t_1 \mathbf{v\check{c}era}_{tt_1} \ldots (o\tau)$
`frequency TIL:` $\mathbf{Onc} \ldots ((o(o\tau))\pi)_\omega$
`verbal object TIL:` $x_1/(o(o\pi)(o\pi)) = \mathbf{p\check{r}ij\acute{i}t}_w \ldots o$
`clause TIL:` $\lambda w_1 \lambda t_2 \Big[\mathbf{P}_{t_2}, \Big[\mathbf{Onc}_{w_1}, \lambda w_3 \lambda t_4 (\exists x_5)\Big($

$\big[\mathbf{Does}_{w_3t_4}, Petr, [\mathbf{Perf}_{w_3}, x_5]\big] \;\wedge\; x_5 = \mathbf{p\check{r}ij\acute{i}t}_{w_3}\Big)\Big], \lambda t_7 \mathbf{v\check{c}era}_{t_2t_7}\Big] \ldots \pi$

and

Markéta telefonovala babičce (Markéta phoned her grandmother):
`time span TIL:` $\mathbf{Anytime} \ldots (o\tau)$
`frequency TIL:` $\mathbf{Onc} \ldots ((o(o\tau))\pi)_\omega$
`verbal object TIL:` $x_1/(o(o\pi)(o\pi)) = [\mathbf{telefonovat}, i_2]_w \wedge$
$\wedge\, [\mathbf{babi\check{c}ka}_{wt}, i_2] \ldots o$
`clause TIL:` $\lambda w_1 \lambda t_2 \Big[\mathbf{P}_{t_2}, \Big[\mathbf{Onc}_{w_1}, \lambda w_3 \lambda t_4 (\exists x_5)(\exists i_6)(\exists i_7)\Big($

$\big[\mathbf{Does}_{w_3t_4}, Mark\acute{e}ta, [\mathbf{Perf}_{w_3}, x_5]\big] \;\wedge\; [\mathbf{babi\check{c}ka}_{w_3t_4}, i_6] \;\wedge$

$$\wedge\, x_5 = [\textbf{telefonovat}, i_6]_{w_3}\Big)\Big], \textbf{Anytime}\Big] \ldots \pi$$

When these two clauses are joined together, the corresponding sentence schema is looked up in the lexicon:

```
sentence_rule_schema: schema = S1 'když' S2
reading #1: lwt(tense_temp(awt(#1),awt(#2)))
```

The first reading corresponds to the *when* translation of 'když', i.e. the sentence is analysed as adverbial temporal clause where we do not ascribe the generating of a time class to the conjunction, but rather use the clause's construction directly as a generator of a collection of the time moments where the clause's extension is True. The subordinate clause is used as a generator of the reference time span – a characteristic function of a class of time moments ($\lambda t_0[\ldots]$).

Thus the original main clauses *reference time span* $\lambda t_7 \textbf{včera}_{tt_7}$ is replaced with

$$\lambda t_1 \Bigg(\textbf{včera}_{tt_2} \;\wedge\; \Bigg[\textbf{P}_t, \Big[\textbf{Onc}_w, \lambda w_5 \lambda t_6 (\exists x_7)(\exists i_8)(\exists i_9)\Big($$

$$\big[\textbf{Does}_{w_5 t_6}, \textit{Markéta}, \quad \ldots \quad \Big)\Big], \lambda t_7 (t_7 = t_1)\Bigg]\Bigg) \ldots ((o\tau)\tau)$$

The resulting analysis of the whole sentence in the first (temporal) reading looks like

$$\lambda w_1 \lambda t_2 \Bigg[\textbf{P}_{t_2}, \Big[\textbf{Onc}_{w_1}, \lambda w_5 \lambda t_6 (\exists x_7)\Big(\big[\textbf{Does}_{w_5 t_6}, \textit{Petr}, [\textbf{Perf}_{w_5}, x_7]\big] \;\wedge$$

$$\wedge\, x_7 = \textbf{přijít}_{w_5}\Big)\Big], \lambda t_9 \Bigg(\textbf{včera}_{t_2 t_9} \;\wedge\; \Bigg[\textbf{P}_{t_2}, \Big[\textbf{Onc}_{w_1},$$

$$\lambda w_{13} \lambda t_{14} (\exists x_{15})(\exists i_{16})\Big(\big[\textbf{Does}_{w_{13} t_{14}}, \textit{Markéta}, [\textbf{Perf}_{w_{13}}, x_{15}]\big] \wedge [\textbf{babička}_{w_{13} t_{14}}, i_{16}] \;\wedge$$

$$x_{15} = [\textbf{telefonovat}, i_{16}]_{w_{13}}\Big)\Big], \lambda t_{18} (t_{18} = t_9)\Bigg]\Bigg)\Bigg] \ldots \pi$$

The second reading of the sentence with 'když' as *if* (*then*) is a sort of *causal adverbial* clauses. In the logical analysis it represents a certain relation between the propositions denoted by the main and causal clauses. Specifically, *if–then* as a *conditional* clause determines the condition (real or unreal) under which the main clause's assertion holds.

In analogy with the analysis of the causal adverbial phrases, we let the clause's conjunction to denote the appropriate relation between the two propositions. The second reading of the example sentence is looked up as

```
reading #2: lwt([awt(0(když/(((o((ot)w)((ot)w))t)w))),#1,#2])
Schema trivialization TIL: když...(oππ)τω
```

The whole sentence with the (in this situation semantically improper) reading as a conditional statement is analysed as

$$\lambda w_1 \lambda t_2 \Bigg[\mathbf{kdy\check{z}}_{w_1 t_2}, \lambda w_3 \lambda t_4 \Bigg[\mathbf{P}_{t_4}, \Bigg[\mathbf{Onc}_{w_3}, \lambda w_5 \lambda t_6 (\exists x_7) \Big(
[\mathbf{Does}_{w_5 t_6}, Petr, [\mathbf{Perf}_{w_5}, x_7]] \ \wedge \ x_7 = \mathbf{p\check{r}ij\acute{i}t}_{w_5} \Big) \Bigg], \lambda t_9 \mathbf{v\check{c}era}_{t_4 t_9} \Bigg],
\lambda w_{10} \lambda t_{11} \Bigg[\mathbf{P}_{t_{11}}, \Bigg[\mathbf{Onc}_{w_{10}}, \lambda w_{12} \lambda t_{13} (\exists x_{14})(\exists i_{15}) \Big(
[\mathbf{Does}_{w_{12} t_{13}}, Mark\acute{e}ta, [\mathbf{Perf}_{w_{12}}, x_{14}]] \ \wedge \ [\mathbf{babi\check{c}ka}_{w_{12} t_{13}}, i_{15}] \ \wedge
\wedge \ x_{14} = [\mathbf{telefonovat}, i_{15}]_{w_{12}} \Big) \Bigg], \mathbf{Anytime} \Bigg] \Bigg] \ldots \pi$$

The conjunction of some causal clauses could be also translated into their logical equivalents (with the help of $\Rightarrow$, $\neg$, $\wedge$ or $\vee$). Since this process may bring some inaccuracies into the analysis, we prefer in the NTA, at least at the moment, to keep the "causal" functions in one to one relation to the actual NL expressions that denote them. However, nothing hinders us from describing the implications between propositions in all respects by means of rules of the inference mechanism.

4 Conclusions

In the paper, we have described the implementation of the complex part of the Normal Translation Algorithm in Transparent Intensional Logic that corresponds to building complex sentences. The resulting logical analysis allows to capture any natural combination of the temporal characteristics of the combined propositions.

We have shown the data structures and lexicons used for possibly ambiguous processing of the semantics of complex sentence based on the conjunctions involved. A special treatment is payed to temporal clauses where their characteristics allow to generate the intensional chronology of the original proposition.

The resulting logical analyses form a quality input for any complex knowledge representation and reasoning system.

Acknowledgements This work has been partly supported by the Ministry of Education of CR within the LINDAT-Clarin project LM2010013, by EC FP7 project ICT-248307 and by the Czech Science Foundation under the project P401/10/0792.

References

1. Harrison, J.: Handbook of practical logic and automated reasoning. Cambridge University Press (2009)
2. Duží, M., Jespersen, B., Materna, P.: Procedural Semantics for Hyperintensional Logic. Foundations and Applications of Transparent Intensional Logic. Volume 17 of Logic, Epistemology and the Unity of Science. Springer, Berlin (2010)
3. Tichý, P.: The Foundations of Frege's Logic. de Gruyter, Berlin, New York (1988)
4. Horák, A., Kadlec, V.: New Meta-grammar Constructs in Czech Language Parser synt. In: Lecture Notes in Artificial Intelligence, Proceedings of Text, Speech and Dialogue 2005, Karlovy Vary, Czech Republic, Springer-Verlag (2005) 85–92
5. Horák, A.: Computer Processing of Czech Syntax and Semantics. Librix.eu, Brno, Czech Republic (2008)
6. Horák, A.: The Normal Translation Algorithm in Transparent Intensional Logic for Czech. PhD thesis, Faculty of Informatics, Masaryk University, Brno (2002)
7. Tichý, P.: The semantics of episodic verbs. Theoretical Linguistics **7** (1980) 264–296
8. Tichý, P.: The logic of temporal discourse. Linguistics and Philosophy **3** (1980) 343–369