Verb Valency Semantic Representation for Deep Linguistic Processing

Aleš Horák[1], Karel Pala[1], Marie Duží[2], Pavel Materna[1]

1: Faculty of Informatics, Masaryk University
Botanicka 68a
602 00 Brno
Czech Republic
{hales,pala}@fi.muni.cz

2: VSB-Technical University of Ostrava
17.listopadu 15
708 33 Ostrava-Poruba
Czech Republic
marie.duzi@vsb.cz

Abstract

In the paper, we describe methods for exploitation of a new lexical database of valency frames (VerbaLex) in relation to Transparent Intensional Logic (TIL). We present a detailed description of the Complex Valency Frames (CVF) as they appear in VerbaLex including basic ontology of the VerbaLex semantic roles.

TIL is a typed logical system developed for natural language semantic representation using TIL logical forms known as constructions. TIL is well suited to handle the difficult language phenomena such as temporal relations, intensionality and propositional attitudes. Here we make use of the long-term development of the Normal Translation Algorithm aimed at automatic translation of natural language sentences into TIL constructions.

We examine the relations between CVFs and TIL constructions of predicate-argument structures and discuss the procedure of automatic acquisition of the verbal object constructions. The exploitation of CVFs in the syntactic parsing is also briefly mentioned.

1 Introduction

In the paper we propose a method to integrate the logical analysis of sentences with the linguistic approach to semantics, exploiting the complex valency frames (CVFs) in the VerbaLex verb valency lexicon, see (Hlaváčková, Horák, Kadlec 2006). To this end we first present a brief survey of the logic we are going to use, namely Transparent Intensional Logic (TIL), which was originated by P. Tichý (Tichý 1988). Theoretical aspects of TIL were further developed in particular by P. Materna (Materna 1998) and also by co-authors of this paper (see, Materna, Duží 2005, Horák 2002). A question may be asked why we do not exploit first order predicate logic (PL1) where some of the presented problems have already been explored and PL1 has been used to represent logical forms. It is a well established fact that PL1 is not able to handle systematically the phenomena like propositional verbs (which, of course, appear in our valency frames), grammatical tenses and modalities (modal verbs and modal particles in natural language). On the other hand, since TIL works with types these problems either do not arise or they can be solved in an intuitive way (see Tichý 1988).

In the second linguistic section we present CVFs by means of which the semantics of verbs in natural language such as Czech or English can be described.

In Section 3 we show how CVFs describe the surface valencies of verbs (i.e. their respective morphological cases in Czech) as well as the semantics of their predicate-argument structure. Concerning the latter we make use of the deep semantic roles expressed by two-level labels based partly on the Top Ontology (EuroWordNet) and partly on the selected literals from Princeton WordNet.

Since so far these two ways of description, namely the logical and linguistic one, have been treated separately, the task we set is to propose a method

Proceedings of the 5th Workshop on Important Unresolved Matters, pages 97–104,
Ann Arbor, June 2005.

of their interrelation and coordination. Needless to say that both ways of description of verb semantics are useful.

Hence we are going to show how to combine a logical description using mostly terms like types, individuals, classes, relations, propositions, or, in general, constructions of these entities, with the linguistic framework capturing the idiosyncratic semantic features of the verbs such as SUBS(liquid:1) or AG(person:1|animal:1).

In Section 4 we adduce an example of the analysis of selected English and Czech verbs for which the above mentioned integration has been proposed.

2 Basics of Transparent Intensional Logic

In this Section we provide an introductory explanation of the main notions of Transparent Intensional Logic (TIL). For exact definitions and details see, e.g., Tichý (1988), Tichý (2004), Materna (1998), Materna (2004) and Materna, Duží (2005). TIL approach to knowledge representation can be characterised as the 'top-down approach'. TIL 'generalises to the hardest case' and obtains the 'less hard cases' by lifting various restrictions that apply only higher up. This way of proceeding is opposite to how semantic theories tend to be built up. The standard approach (e.g. predicate logic) consists in beginning with atomic sentences, then proceeding to molecular sentences formed by means of truth-functional connectives or by quantifiers, and from there to sentences containing modal operators and, finally, attitudinal operators.

Thus, to use a simple case for illustration, once a vocabulary and rules of formation have been laid down, semantics gets off the ground by analysing an atomic sentence as follows:

(1) "Charles is happy": Fa

And further upwards:

(2) "Charles is happy, and Thelma is grumpy": $Fa \wedge Gb$

(3) "Somebody is happy": $\exists x\ (Fx)$

(4) "Possibly, Charles is happy": $\Diamond\ (Fa)$

(5) "Thelma believes that Charles is happy": $\mathrm{B}b\ (Fa)$.

In non-hyperintensional (i.e., non-procedural) theories of formal semantics, attitudinal operators are swallowed by the modal ones. But when they are not, we have three levels of granularity: the coarse level of truth-values, the fine-grained level of truth-conditions (propositions, truth-values-in-intension), and the very fine-grained level of hyper-propositions, i.e., *constructions* of propositions. TIL operates with these three levels of granularity. We start out by analysing sentences from the uppermost end, furnishing them with a hyperintensional[1] semantics, and working our way downwards, furnishing even the lowest-end sentences (and other empirical expressions) with a hyperintensional semantics. That is, the sense of a sentence such as "Charles is happy" is a hyper-proposition, namely the *construction of the denoted proposition* (i.e., the *instruction* how to evaluate the truth-conditions of the sentence in any state of affairs).

When assigning a construction to an expression as its meaning, we specify a *procedural know-how*, which must not be confused with the respective *performancy know-how*. Distinguishing performatory know-how from procedural know-how, the latter could be characterised "that a knower x knows how A is done in the sense that x can spell out instructions for doing A." For instance, to *know* what Goldbach Conjecture *means* is to understand the instruction to find whether 'all positive even integers ≥ 4 can be expressed as the sum of two primes'. It does not include either actually finding out (whether it is true or not by following a procedure or by luck) or possessing the skill to do so.[2]

Furthermore, the sentence "Charles is happy" is an 'intensional context', in the sense that its logical analysis must involve reference to empirical parameters, in this case both possible worlds and instants of time. Charles is only *contingently* happy; i.e., he is only happy at some worlds and only sometimes. The other reason is because the *analysans* must be capable of figuring as an argument for functions whose domain are propositions rather than truth-values. Construing 'Fa' as a name of a truth-value works only in the case of (1), (2) and (3). It won't work in (5), since truth-values are not the sort of thing that can be

[1] The term 'hyperintensional' has been introduced by Max Cresswell in Cresswell (1975). See also Cresswell (1985).

[2] For details on TIL handling knowledge see Duží, Jespersen, Müller (2005).

believed. Nor will it work in (4), since truth-values are not the sort of thing that can be possible.

Constructions are procedures, or instructions, specifying how to arrive at less-structured entities. Being procedures, constructions are structured from the algorithmic point of view, unlike set-theoretical objects. The TIL 'language of constructions' is a modified hyper-intensional version of the typed λ-calculus, where Montague-like λ-terms denote, not the functions constructed, but the constructions themselves. Constructions *qua* procedures operate on input objects (of any type, even on constructions of any order) and yield as output (or, in well defined cases fail to yield) objects of any type; in this way constructions construct *partial* functions, and functions, rather than relations, are basic objects of our ontology.

By claiming that constructions are algorithmically structured, we mean the following: a construction *C* — being an instruction — consists of particular steps, i.e., sub-instructions (or, *constituents*) that have to be executed in order to execute *C*. The concrete/abstract objects an instruction operates on are not its constituents, they are just mentioned. Hence objects have to be supplied by another (albeit trivial) construction. The constructions themselves may also be only mentioned: therefore one should not conflate *using* constructions as constituents of composed constructions and *mentioning* constructions that enter as input into composed constructions, so we have to strictly distinguish between *using* and *mentioning* constructions. Just briefly: Mentioning is, in principle, achieved by using atomic constructions. A construction is atomic if it is a procedure that does not contain any other construction as a used subconstruction (a constituent). There are two atomic constructions that supply objects (of any type) on which complex constructions operate: *variables* and *trivializations.*

Variables are constructions that construct an object dependently on *valuation*: they *v-construct,* where *v* is the parameter of valuations. When X is an object (including constructions) of any type, the *Trivialization* of X, denoted $^{0}\boldsymbol{X}$, constructs X without the mediation of any other construction. ^{0}X is the atomic concept of X: it is the primitive, non-perspectival mode of presentation of X.

There are three *compound* constructions, which consist of other constructions: *Composition*, *Closure* and Double Execution. Composition $[\boldsymbol{X}\ \boldsymbol{Y}_1 \ldots \boldsymbol{Y}_m]$ is the procedure of applying a function *f* *v*-constructed by *X* to an argument A *v*-constructed by $Y_1,\ldots,Y_m$, i.e., the instruction to apply *f* to A to obtain the value (if any) of *f* at A. Closure $[\boldsymbol{\lambda x_1 \ldots x_m}\ \boldsymbol{Y}]$ is the procedure of constructing a function by abstracting over variables, i.e., the instruction to do so. Finally, higher-order construction *X* can be used twice over as a constituent of a composed construction. This is achieved by the fifth construction called *Double Execution* $^{2}\boldsymbol{X}$.

TIL constructions, as well as the entities they construct, all receive a type. On the ground level of the type-hierarchy, there are entities unstructured from the algorithmic point of view belonging to a *type of order 1*. Given a so-called *epistemic (or 'objectual')* ***base*** of ***atomic types*** (**ο**-truth values, **ι**-individuals, **τ**-time moments / real numbers, **ω**-possible worlds), mereological complexity is increased by the induction rule for forming partial functions: where $\alpha, \beta_1,\ldots,\beta_n$ are types of order 1, the set of partial mappings from $\beta_1 \times\ldots\times \beta_n$ to α, denoted $(\alpha\beta_1\ldots\beta_n)$, is a type of order 1 as well. Constructions that construct entities of order 1 are *constructions of order 1*. They belong to a *type of order 2*, denoted by $*_1$. Inductively we define *type of order n,* $*_n$.

TIL is specific in a precise solution for intensions as non-empirical objects of the real world. Intensions are qualified as functions of a type ((ατ)ω), i.e., functions from possible worlds to *chronologies* of the type α (in symbols: $\alpha_{\tau\omega}$), where a chronology is a function of type (ατ). Some important kinds of intensions are:

Propositions, type $o_{\tau\omega}$ (shortened as π). They are denoted by empirical (declarative) sentences.

Properties of members of a type α, or simply *α-properties,* type $(o\alpha)_{\tau\omega}$.[3] General terms (some substantives, intransitive verbs) denote properties, mostly of individuals.

Relations-in-intension, type $(o\beta_1\ldots\beta_m)_{\tau\omega}$. For example transitive empirical verbs, also attitudinal verbs denote these relations. Omitting $_{\tau\omega}$ we get the type $(o\beta_1\ldots\beta_m)$ of *relations-in-extension* (to be met mainly in mathematics).

[3] Collections, sets, classes of 'α-objects' are members of type (oα); TIL handles classes (subsets of a type) as characteristic functions. Similarly relations (-in-extension) are of type(s) $(o\beta_1\ldots\beta_m)$.

α-roles or offices, type $\alpha_{\tau\omega}$, where $\alpha \neq (o\beta)$. Frequently $\iota_{\tau\omega}$ (an individual office). Often denoted by concatenation of a superlative and a noun ("*the highest mountain"*). Individual roles correspond to what Church calls an "*individual concept"*.

3 The Complex Valency Frames

Valency frames have been built in several projects (VALLEX for Czech PDT (Žabokrtský 2005) or VerbNet (Kipper et al 2006)). Motivation for the VerbaLex project came from comparing Czech WordNet verb frames with VALLEX. The main goal of VerbaLex is an automatic processing of verb phrases exploiting explicit links to Princeton WordNet. The complex valency frames we are working with can be characterized as data structures (tree graphs) describing predicate-argument structure of a verb which contains the verb itself and the arguments determined by the verb meaning (their number usually varies from 1-5). The argument structure also displays the semantic preferences on the arguments. On the syntactic (surface) level the arguments are most frequently expressed as noun or pronominal groups in one of the seven cases (in Czech) and also as prepositional cases or adverbials.
An example of a complex valency frame for the verb *zabít (kill)* looks like:
usmrtit:1/zabít:1/dostat:11 (kill:1)
-frame: AG<person:1|animal:1>$_{\text{who\_nom}}^{\text{obl}}$
VERB$^{\text{obl}}$
PAT<person:1|animal:1>$_{\text{whom\_acc}}^{\text{obl}}$
INS<instrument:1>$_{\text{with\_what\_ins}}^{\text{opt}}$
-example: vrah zabil svou oběť nožem (A murderer has killed the victim with a knife).
-synonym:
-use: prim
More examples of CVFs for some selected verbs can be found below in Section 4.
The semantics of the arguments is typically labeled as belonging to a given semantic role (or deep case), which represents a general role plus subcategorization features (or selectional restrictions). Thus valency frames in Verbalex include information about:

1. the syntactic (surface) information about the syntactic valencies of a verb, i.e. what **morphological cases** (direct and prepositional ones in highly inflected languages such as Czech) are associated with (required by) a particular verb, and also **adverbials**,
2. **semantic roles** (deep cases) that represent the integration of the general labels with subcategorization features (or selectional restrictions) required by the meaning of the verb.

The inventory of the semantic roles is partly inspired by the Top Ontology and Base Concepts as they have been defined within EuroWordNet project. Thus we work with the general or 'large' roles like AG, ART(IFACT), SUBS(TANCE), PART, CAUSE, OBJ(ECT) (natural object), INFO(RMATION), FOOD, GARMENT, VEHICLE and others. They are combined with the literals from Princeton WordNet 2.0 where literals represent subcategorization features allowing us to climb down the hypero/hyponymical trees to the individual lexical units. For example, we have AG(person:1|animal:1) or SUBS(liquid:1) that can be used within the individual CVFs.
The verb entries are linked to the Czech and Princeton WordNet 2.0, i.e. they are organized around the respective lemma in synsets with numbered senses.
The Czech lexical resource being now developed is then a list of Czech CVFs – this work is going on within the Verbalex project at FI MU (Hlaváčková, Horák, 2005). Verbalex now contains approx. 11000 verb literals organized in synsets. The current goal is to enlarge the lexicon to 15 000 verbs.
The inventory of the semantic roles we work with clearly represents a sort of ontology which tries to cover word stock of Czech verbs and can be used as a base for a semantic classification and subclassification of the verbs. The ontologies represent theoretical constructs designed from the „top" and as such they are not directly based on the empirical evidence, i.e. corpus data. Thus there is a need to confront the ontologies and the inventories of the semantic roles that can be derived from them with the corpus data and see how well they can correspond to them. For this purpose we are experimenting with the corpus data obtained from the Word Sketch Engine (Kilgarriff, Rychlý, Smrž, Tugwell 2006).

4 Logical Analysis Using CVFs

In this section we describe the translation of VerbaLex CVFs into a verb phrase, which is a core of a sentence logical analysis.

TIL comes with a dissociation of significant verbs into two groups according to the classification of their meaning:

1. *by attributive verbs* we ascribe qualities or properties to objects. Attributive verbs are typically expressed by the respective form of the verb 'to be' combined with an expression denoting a property; examples: 'to be red' or 'to be mellow' or with a general substantive like 'to be a traitor', 'to be a tree'.
2. *episodic verbs*, on the other hand, specify actions performed by a subject.

An episodic verb does not describe its subject's state in any moment of time, it rather describes an episode of doing something at the certain time moment (and necessarily some time before that moment plus the expectation that it will last also in the next few moments, at least). *TIL* provides a complex handling of episodic verbs including the verb tense, aspect (perfective/imperfective) or active/passive state. All these features are concentrated around the so called *verbal object*, the construction of which (i.e., the meaning of a particular verb phrase) is the application of (the construction of) the verb to (the constructions of) the verb's arguments.

Since the analysis of attributive verbs is usually quite simple, we will concentrate in the following text on the examples of selected episodic verbs from VerbaLex and their logical analysis using the complex valency frames.

The TIL type of episodic verbal objects is $(o(o\pi)(o\pi))_{\omega}$, where π is the type of propositions ($o_{\tau\omega}$). See (Horák 2002, pp. 64-73) and (Tichý 1980) for detailed explanation. Our analysis is driven by a linguistic (syntactic) context that signals the *semantic fact that there is always a function involved here,* so that we have to ascribe types to its arguments and value.

4.1 Examples of Logical Analysis

We have chosen cca 10 verbs with their verb frames from VerbaLex and we will use them as examples of the algorithm for determining the verb type in the TIL logical analysis procedure.

dát (give)

dát:2 / dávat:2 / darovat:1 / věnovat:1 (give:8, gift:2, present:7)

-frame: DON<organization:1>$_{what\_nom}^{obl}$ VERB^{obl} OBJ<object:1>$_{what\_acc}^{obl}$ BEN<person:1>$_{to\_whom\_dat}^{obl}$

-example: firma věnovala zaměstnancům nová auta (a company gave new cars to the employees)

-use: prim

The verb arguments in this frame are: *who, to whom, what* (all obligatory) with (at least) two options: a) *to whom* is an individual, b) *to whom* is a class of individuals. The respective verb types are ad a): $((o(o\pi)(o\pi))_{\omega}\iota\iota\iota)$,

ad b): $((o(o\pi)(o\pi))_{\omega}\iota(o\iota)\iota)$.

For example *to whom* = to the employees of a given institution. To be an employee of the institution XY is a property, say Z / $(o\iota)_{\tau\omega}$. So "The company gave to the employees of XY…", not taking into account grammatical tenses and omitting trivializations we get $\lambda w\lambda t$ [Give$_{wt}$ XY Z$_{wt}$ etc.] (XY has the type ι here, being a collective rather than a class.)

With this example, we can show that CVFs are used not only for determining the verbal object type, but also for stating additional *prerequisities* (necessary conditions) for the sentence constituents. The full analysis using the verb frame above thus contains, except the verb phrase part, the conditions saying that "*X gives Y to Z* $\wedge$ *organization(X)* $\wedge$ *object(Y)* $\wedge$ *person(Z)*". The predicates *organization*, *object* and *person* here represent the properties denoted by the corresponding terms in the wordnet hypero-hyponymical hierarchy.

dát:15 / dávat:15 / nabídnout:3 / nabízet:3 (give:37)

-frame: AG<person:1>$_{who\_nom}^{obl}$ VERB^{obl} ABS<abstraction:1>$_{what\_acc}^{obl}$ REC<person:1>$_{to\_whom\_dat}^{obl}$

-example: dal jí své slovo (he gave her his word)

-example: nabídl jí své srdce (he offered her his heart)

-use: fig

Here we have an idiom ("to give word"), which corresponds to an (episodic) relation between two individuals. Thus the type of the verb is $((o(o\pi)(o\pi))_{\omega}\iota\iota)$, the second ι corresponds to *to whom*.

bránit (prevent)
bránit:1 / zabránit:2 / zabraňovat:2 / zamezit:2 / zamezovat:2 (prevent:2, keep:4)
-frame: AG<person:1>$_{who\_nom}^{obl}$ VERB^{obl} PAT<person:1>$_{to\_whom\_dat}^{obl}$ ACT<act:1>$_{inf}^{obl}$
-example: zabránila mu uhodit syna (she prevented him from hitting the son)
-use: prim

bránit:1 / zabránit:2 / zabraňovat:2 / zamezit:2 / zamezovat:2 (prevent:2, keep:4)
-frame: AG<institution:1>$_{what\_nom}^{obl}$ VERB^{obl} PAT<person:1>$_{to\_whom\_dat}^{obl}$
ACT<act:2>$_{in\_what\_loc}^{opt}$
-example: policie mu zabránila v cestě do zahraničí (police prevented him from going abroad)
-use: prim

Here, arguments of the verb correspond to the phrases *who, to whom, in (from)*. The third argument has the type of an activity given, of course, by an episodic verb *hit the son, travel abroad* (the substantive form *travelling abroad* can be construed as that activity). The type of the verb is $((o(o\pi)(o\pi))_{\omega}\iota\iota((o(o\pi)(o\pi))_{\omega}))$.

říct (say)
říct:1 / říkat:1 / říci:1 / říkat:1 / pravit:1 (say:6)
-frame: AG<person:1>$_{who\_nom}^{obl}$ VERB^{obl} COM<speech act:1>$_{what\_acc,that,dsp}^{obl}$ ADR<person:1>$_{to\_whom\_dat}^{opt}$
-example: říct kolegovi dobrý den (say hello to a colleague)
-example: řekl, že to platí (he said that it holds)
-example: pravil: "Dobrý den" (he said: “Good day”)
-use: prim

The case questions for the corresponding arguments of the verb *říct* are *a) who, what*$_1$, *b) who, what*$_2$, *c) who, to whom, what*$_1$, and *d) who, to whom, what*$_2$. Examples of instantiated sentences can be a) *Charles says „Hello“*, b) *Charles says that he is ill,* c) *Charles says to his colleague “Hello”*, or d) *Charles says to his colleague that he is ill*.

The quotation context (*ad* a), c)) is normally impossible to type. Unless we want to go into some deep analyses we can ascribe to any quoted expression the type of individual. The relation to and unquoted subordinate clause is analysed as a general construction of type $*_n$. The resulting types of verbs are then
a) $((o(o\pi)(o\pi))_{\omega}\iota\iota)$,
b) $((o(o\pi)(o\pi))_{\omega}\iota *_n)$,
c) $((o(o\pi)(o\pi))_{\omega}\iota\iota\iota)$,
d) $((o(o\pi)(o\pi))_{\omega}\iota\iota *_n)$.

brečet$_1$ (cry) because of something, for something
brečet:1 / plakat:1 (cry:2, weep:1)
-frame: AG<person:1>$_{who\_nom}^{obl}$ VERB^{obl} CAUSE<cause:4>$_{due+to+what\_dat,over+what\_ins,for+what\_acc}^{obl}$
-example: brečela kvůli zničeným šatům (she cried for spoiled clothes)
-example: plakal nad svou chudobou (he cried over his poverty)
-example: plakal pro své hříchy (he cried for his sins)
-use: prim

brečet$_2$ (cry) for somebody
brečet:1 / plakat:1 (cry:2, weep:1)
-frame: AG<person:1>$_{who\_nom}^{obl}$ VERB^{obl} ENT<person:1>$_{for+whom\_acc}^{obl}$
-example: plakala pro milého (she cried for her boy)
-use: prim

If I cry *because of, for* etc., then the role of causing is played by this *because of*. Crying is an episodic verb, whereas *because of* etc. is a relation between propositions, often between events. We have therefore *because of* / $(o\pi\pi)_{\tau\omega}$, where the first $\pi(=o_{\tau\omega})$ belongs to the proposition denoted, e.g., by *clothes have been spoiled* or that *the respective individual is poor, sinful* etc., and the second π to the proposition *that the respective individual cries*.
In case of *to cry for somebody* the respective type is again a “relation” $((o(o\pi)(o\pi))_{\omega}\iota\iota)$, although this *for* hides some cause, which is, however, not mentioned.
With this verb, we will describe the analysis of verb entailment handling in TIL. If we analyse a general case of the above mentioned meanings of cry (cry$_1$-because of something, cry$_2$-for

somebody) simply *to cry,* (*He cries all the time*). This verb's type is a verbal object without arguments, $(o(o\pi)(o\pi))_{\omega}$. In addition to this the following rule holds: *If X cries because of… or X cries for…, then X cries.* In this way the semantic dependence between the three cases of crying is given; otherwise we would not be able to detect this connection, e.g. between **brečet$_1$** and **brečet$_2$**.

absolvovat (undergo)

absolvovat:2 / prožít:1 / prožívat:1 (experience:1, undergo:2, see:21, go through:1)

-frame: AG<person:1>$_{who\_nom}^{obl}$ VERB^{obl} EVEN<experience:3>$_{what\_acc}^{obl}$ LOC<location:1>$_{in\_what\_loc}^{opt}$

-example: absolvoval vyšetření na psychiatrické klinice (he went through investigation in a psychiatric clinic)

-use: prim

In general it is an episodic relation to an event (type π)[4], so the type is $((o(o\pi)(o\pi))_{\omega}\iota\pi)$. In some cases we may also use a relation to an episode (specific class of events, type $(o\pi)$), then the type is $((o(o\pi)(o\pi))_{\omega}\iota(o\pi))$, and *investigation in a clinic* has to be defined as a sequence of events.

akceptovat (accept)

akceptovat:3 / přijmout:6 / přijímat:6 (accept:4)

-frame: AG<person:1|social group:1>$_{who\_nom}^{obl}$ VERB^{obl} STATE<state:4>|EVEN<event:1>|INFO<info:1>$_{what\_acc}^{obl}$

-example: akceptujeme jeho povahu (we accept his character)

-example: lidé přijali nový zákon s nadšením (people accepted new law with enthusiasm)

-use: prim

We can accept nearly anything. Here we meet the problem of type-theoretical polymorphism, which is handled here as a type *scheme* $((o(o\pi)(o\pi))_{\omega}\iota\alpha)$, for an arbitrary type α. A quintessence of such a polymorphism: *think on (about)* — one can think of an object of any kind.

učit (teach)

naučit:1 / učit:2 / vyučovat:1 (teach:1, learn:5, instruct:1)

-frame: AG<person:1>$_{who\_nom}^{obl}$ VERB^{obl} PAT<person:1>$_{whom\_acc}^{opt}$ KNOW<subject:3>$_{what\_acc,to\_what\_dat}^{obl}$

-example: naučil dítě abecedu (he educated a children in the alphabet)

-example: učí studenty matematiku (he teaches mathematics for students)

-example: vyučuje dějepisu (he/she teaches history)

-use: prim

If understood as in "*What does (s)he live off? (S)he teaches.*" it is the case of *cry$_3$* (see above). *To teach* understood as in "*He teaches history, maths*", etc., the analysis depends on which type is given to the school subjects, disciplines. One possibility is to analyse them as properties of a set of propositions, $(o(o\pi))_{\tau\omega}$. Then *to teach* receives the type $((o(o\pi)(o\pi))_{\omega}\iota(o(o\pi))_{\tau\omega})$. If "*teaches alphabet*" is the case then we have to decide what we mean by *alphabet*. Here the point is to *teach (learn)* to associate symbols and sounds (phonemes?), so the respective type of alphabet is $(\alpha\beta)$, where α is the type of symbols, β the type of sounds. In the analysis of "*to educate somebody in something*" the verb takes an individual as its additional argument: $((o(o\pi)(o\pi))_{\omega}\iota\iota\alpha)$, where α is the type of the discipline.

In all the examples, we have displayed the relations between the two-level semantic roles used in the VerbaLex verb frames and the resulting logical analysis types of the verbal object as the main part of the clause's logical construction. The algorithmisation of this procedure uses a list of all roles used in the lexicon (there are about 200 roles used) with the corresponding (ambiguous) logical types of the constituents. In this way we can form a basic skeleton of the automatic translation of text to logical constructions.

5 Conclusions

The paper presented a first outline of comparison and integration of the two approaches, namely logical and linguistic, to the semantics of verbs in a natural language (English and Czech). We are aware that this work is still in a great progress and the results so presented rather fragmentary. Still, we are convinced that the research project we aim at is a relevant contribution to the semantics of natural language.

[4] see (Horák 2002, p. 65) and (Tichý 1980).

We have shown that pursuing such a research is reasonable and comes up with a new viewpoint to the meaning of verbs. In this way we extend our knowledge in the important way. Actually, we are dealing with two deep levels of the meaning description and a question may be asked which one is deeper and better. Our answer is, do not contrast the two levels, and make use of both of them. In this way we believe to integrate them into one compact whole and perhaps obtain a unique data structure. The results of the presented research can be immediately applied in the area of knowledge representation and in the long-term Normal Translation System project that is being prepared.
We have not tackled the other deep descriptions, such as the method that exploits the tectogramatical level as it is presently applied in PDT (Hajič 2004). This, obviously, is a topic of another paper.

Acknowledgments

This work has been supported by the Academy of Sciences of the Czech Republic, project No. T100300414, by the Ministry of Education of CR within the Center of basic research LC536, by the program 'Information Society' of Czech Academy of Sciences, project No. 1ET101940420 "Logic and Artificial Intelligence for multi-agent systems", and by the Czech Science Foundation under the project 201/05/2781.

References

Cresswell, M.J. (1975): 'Hyperintensional Logic'. *Studia Logica* 34, pp.25-38.

Cresswell, M.J. (1985): *Structured meanings.* MIT Press, Cambridge, Mass.

Duží, M., Jespersen, B., Müller, J. (2005): Epistemic Closure and Inferable Knowledge. *The Logica Yearbook 2004*, ed. L. Běhounek, M. Bílková, Filosofia Prague, pp. 125-140.

Fellbaum, C., editor. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, Massachusetts, London, England.

Hajič, Jan (2004): *Complex Corpus Annotation: The Prague Dependency Treebank*, Jazykovedny Ustav L.Stura, Bratislava, Slovakia, 2004.

Hlaváčková, Dana - Horák, Aleš - Kadlec, Vladimír (2006). Exploitation of the VerbaLex Verb Valency Lexicon in the Syntactic Analysis of Czech. Lecture Notes in Artificial Intelligence, Proceedings of Text, Speech and Dialogue 2006, Berlin, Heidelberg : Springer, 2006.

Horák, Aleš (2002). *The Normal Translation Algorithm in Transparent Intensional Logic for Czech*, Ph.D. Dissertation, Masaryk University, Brno, 2002.

Kilgarriff, Adam - Rychlý, Pavel - Smrž, Pavel - Tugwell, David (2006). *The Sketch Engine*. In Proceedings of the Eleventh EURALEX International Congress. Lorient, France : Universite de Bretagne-Sud, pp. 105-116, 2004.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer (2006): Extensive Classifications of English verbs. *Proceedings of the 12th EURALEX International Congress.* Turin, Italy. September, 2006.

Materna, P. (1998): *Concepts and Objects.* Acta Philosophica Fennica, Vol. 63, Helsinki.

Materna, P. (2004): *Conceptual Systems.* Logos Verlag, Berlin.

Materna, P., Duží, M. (2005): Parmenides Principle. *Philosophia,* vol. 32 (1-4), pp. 155-180.

Tichý, P. (1988): *The Foundations of Frege's Logic*, Berlin, New York: DeGruyter.

Tichý, P. (1980): The Semantics of Episodic Verbs, *Theoretical Linguistics 7*, pp. 263-296, 1980.

Tichý, P. (2004): *Collected Papers in Logic and Philosophy*, V. Svoboda, B. Jespersen, C. Cheyne (eds.), Prague: Filosofia, Czech Academy of Sciences, and Dunedin: University of Otago Press

Žabokrtský, Z. (2005): *Valency Lexicon of Czech Verbs.* Ph.D. Thesis, Faculty of Mathematics and Physics, Charles University in Prague, 2005.