The Ordered-triple Theory of Language: Its History and the Current Context

Aleš Horák and Karel Pala

Natural Language Processing Centre
Faculty of Informatics, Masaryk University
Botanická 68a, 602 00, Brno, Czech Republic
hales@fi.muni.cz, pala@fi.muni.cz

Abstract. In this paper, we recall the historical perspectives of the Ordered-Triple Theory of Language (OTT) whose authors are Materna, Pala and Svoboda. The Ordered-Triple Theory, as the title suggests captures three fundamental components of a language system, i.e. syntax, semantics and pragmatics, and is fully comparable with similar linguistic theories. It became a starting point for further interconnection of logic, linguistics and informatics thanks to the intensive mutual cooperation of Pala and Materna at the newly established Faculty of Informatics from 1995.
We show the subsequent milestones related to OTT and its realisation by means of the transparent intensional logic (TIL) in relation to the natural language processing (primarily Czech).

Key words: ordered-triple theory, theory of language, syntax, semantics, pragmatics, procedural grammar of Czech, transparent intensional logic, TIL

1 Introduction

The authors of the Ordered-triple Theory of Language (further OTT) are Pavel Materna, Karel Pala a Aleš Svoboda who proposed it at FF UJEP[1] during 1976–79 and published in *Brno Studies in English 12* [1] and *13* [2].

In a sense, OTT was a reaction to the two language theories that were influential by this time, particularly to Chomsky's generative approach [3] and the Prague functional generative framework (FGD) by Sgall et al [4]. The main difference is that OTT had not been conceived explicitly as generative but it had allowed to deal with both recognoscative and generative devices (see below). Prague's FGD from the beginning contained semantic component in the form of the tectogrammatical level which was based on a set of actants (semantic roles). However, it did not use any logical formalism in contrast with OTT. The relevant feature of OTT was a consistently grasped semantic component while, for instance, Chomsky's generative grammars were primarily based on syntax.

[1] The Faculty of Arts at the former Jan Evangelista Purkyně University, current Masaryk University, in Brno

Aleš Horák, Pavel Rychlý, Adam Rambousek (Eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2017, pp. 29–36, 2017.

Later, attempts appeared to mate the generative approach with Montague's intensional logic [5,6].

In the following text we would like to characterize OTT briefly and present the main interesting results that have been reached within this framework.

2 The Ordered-triple Theory

The Ordered-triple Theory offers a theoretical framework for a formal natural language description which considers all the basic components of any semiotic system, i.e. it captures relations between language user and real world and relations between language and language user. In other words, the proposed framework consequently takes into account the syntactic, semantic and pragmatic component of language (in Morris' sense [7]) and takes them as a unified system. The assumption is that natural language expressions consist of syntactic, semantic and pragmatic constituents and thus can be described as ordered triples comprising:

⟨semantic component, formal language expression, pragmatic component⟩

2.1 Part I – Semantics

The first part of OTT is mainly about semantics. It is conceived as consequently intensionalistic, thus at the beginning we pay attention to the problems of the extensional approach to semantics in which language expressions denote what we call extensions, i.e. mainly individuals, classes, relations and truth values. Then we give reasons for a different approach that does not suffer from the non intuitive consequences of the extensionalism – extensional analysis does not allow to distinguish empiric sentences from non-empiric ones and understanding from verification. The intensionalistic approach allows to handle also other referential phenomena like individual roles, propositional attitudes, or episodic verbs.

The logical analysis of natural language in OTT relies on the transparent intensional logic (further TIL) in the form of the system, whose author was in 1970s P. Tichý, who after August 1968 left ČSSR,[2] spent short time in UK and then moved to New Zealand where he started to work in Dunedin at the Otago University [8].

We define the basic concepts of the intensional semantics consisting of the epistemic base constituted by four sets: the universe (ι, set of individuals), the set of truth values (o), the set of possible worlds (ω) and the (continuous) set of time moments (τ). The simple type theory is used to produce derived entities and the most typical intensions are given using the operation called *intensional descent*, i.e. the application of a (possible) *world w* and a *time moment t* to arrive from an intension (an $((\alpha\tau)\omega)$-function) to the corresponding extension (an α-object, where α denotes an extensional type).

[2] The Czechoslovak Socialist Republic

Further relevant concepts are introduced: constructions (atoms, applications, abstractions) and the relation between language expression, construction and intension. Also class of what is called language constructions is distinguished among all other constructions. Class of the language constructions is to be understood as a class of the constructions that can be expressed by natural language expressions. In this respect, a grammar of language can be taken as a set of rules enabling to derive constructions reflecting these expressions from the structure of the language expressions: simple examples of such rules are given. Such grammar may contain syntactic rules having the form of context-free rules and semantic rules operating on the output of the grammar and providing formulae of the λ-calculus as a result.

Consequently, it is shown how to extend the sets consisting of the universe, truth values and possible worlds with further sets. Thus the set of the time moments allows to perform a more subtle semantic analysis capturing time characteristics including grammatical tenses. Similarly, the semantics of locational adverbs can be examined if we add to them the set of space points.

The next task is to handle deictic (indexical) expressions as e.g. personal pronouns, and to establish relations between what we call external pragmatics and semantics. Finally, the attention is paid to the semantic relations of expressing, denoting and constructing on one hand and to the pragmatic relations of demonstrating (internal pragmatics) and determining (external pragmatics) on the other.

2.2 Part II – Computer Tools for the OTT

So far, we have characterized the general framework of the OTT, now we would like to deal with its computer application (model). In Chomsky's and Sgall's approaches mentioned above the language levels are used. Within the OTT, we can also have syntactic and semantic component including morphology and, moreover, we are interested in their algorithmic description leading to a computer model for OTT.

In the first version of OTT, a context-free like grammar was used, particularly, it was a set of procedural rules for Czech [9] called a procedural grammar inspired by T. Winograd for English [10]. It was the first and only procedural grammar for Czech implemented in programming language LISP 1.5, containing 34 LISP functions, and tested on the mainframe TESLA 200. It has to be remarked that in 1976 no standard morphological analyzer for Czech existed, therefore a morphological input for the Czech procedural grammar was prepared manually and had a form of a file of word forms with corresponding parts of speech and lists of grammatical features. This "syntactic dictionary" is, in fact, almost identical with the output of the present-day morphological analyzers including our Majka [11,12].

The analyser produced a syntactic structure of a sentence in the form of the labelled tree graph which could serve as an input for the semantic analyser [13] as well as for rules handling attitudes of language users in the internal pragmatics framework (this is missing both in Chomsky's and Prague approaches). The

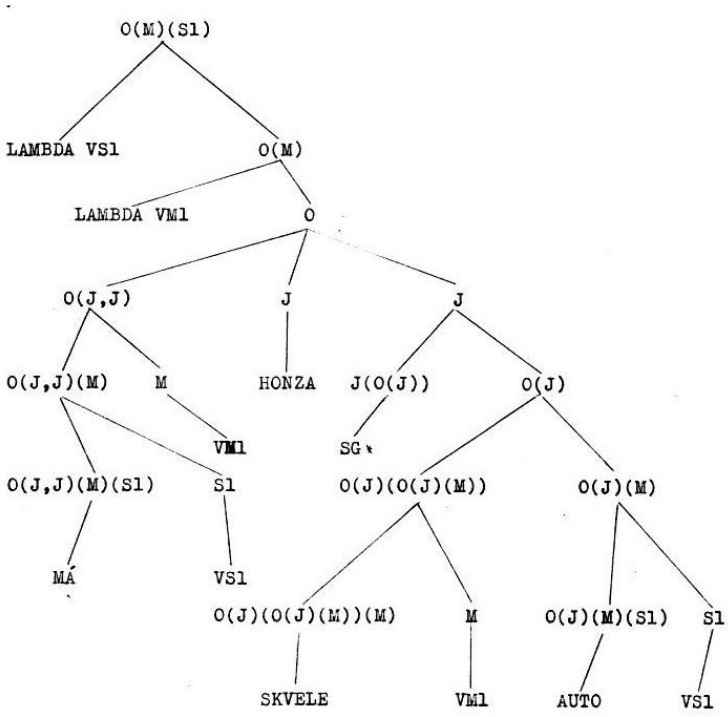

Fig. 1. An example of a semantic tree from [13].

semantic analyser was implemented in LISP 1.5 and tested on TESLA 200 as well. It produced a tree structure representing a λ-calculus formulae obtained from the syntactic tree of the Czech sentence produced by the procedural syntactic analyser (see Figure 1 for an example). λ-calculus formulae are corresponding to the natural language constructions and can be understood as semantic representations of the analyzed natural language (in this case Czech) sentences.

2.3 Other Theoretical Approaches

With contentment, we can say that the mentioned results obtained in OTT (procedural grammar of Czech, semantic analyser, both written in LISP) were in their time fully comparable with the Sgall's FGD framework and with the similar results obtained in the area of transformational grammars. It has to be stressed that OTT has proved itself as a consistent starting point for computer processing of the natural language, particularly Czech, in contrast with American results oriented primarily towards English.

A historical remark [14]: above we have been dealing mainly with the results achieved until 1989. After this A. Svoboda had to move to Opava University,

Materna and Pala finally started to teach at the Faculty of Informatics at Masaryk University (FI MU), where the work on OTT in a sense further continued.

3 Continuation of the OTT at FI MU

The Faculty of Informatics MU was established in 1994, Pala started teaching there in 1995, as well as P. Materna who offered lectures about TIL to the students of informatics in the same year. Thanks to this, Materna got in touch with students who were interested in logical analysis of natural language. One of them was in 1995 A. Horák, who became strongly attracted by the one of the central topic of AI, particularly, by the analysis of the natural language and especially the analysis of meaning. He investigated the problems of the logical and semantic analysis of language in his diploma thesis and also dissertation. The purpose of this work was to make a progress in TIL, which is an important part of the OTT, and continue with the computer implementation enabling to translate standard natural language sentences (Czech, in the first step) into the constructions of the intensional logic. As we have said above, first steps in this direction were presented in [13] and also [15].

However, Horák's research went further and brought a new original result published as the Normal Translation Algorithm (NTA [16]) containing the syntactic analyser Synt [17,18,19] which employs context actions translating syntactic trees of standard Czech sentences to corresponding intensional constructions (expressed as λ-formulae). These results became a base for a further cooperation with P. Materna and later also with M. Duží which led to further development of TIL [20] and the corresponding approaches within several grant projects (GAČR 2005–2007, 2010–2012, 2015–2017), firstly taking place at the FI MU and then also at the Technical University of Ostrava.

The verification of the TIL logical analysis theory included building a large corpus of TIL logical constructions [21] suitable for explicating various language phenomena in common Czech texts. The corpus consists of more than 5,000 sentences that were semi-automatically analysed and translated according to NTA and used for checking by human logicians. An example logical analysis from this corpus is displayed in Figure 2.

3.1 Valency Frames and the OTT

One of the recent results in the NLP Centre is a valency database for Czech language named VerbaLex [22,23]. It contains approx. 10,500 Czech verbs and it is the largest valency database for Czech. Since the verb valency frames represent verbs as predicates with their arguments they can be linked with the linguistic constructions in the TIL [24]. From this point of view we can relate the VerbaLex to the OTT and to exploit it in our further research.

Družice zaznamenaly zrod třetího přechodně trvajícího radiačního pásu Země.

$\lambda w_1 \lambda t_2 [\mathbf{P}_{t_2},$	
$[\mathbf{Onc}_{w_1},$	
$\lambda w_3 \lambda t_4 (\exists x_5)(\exists x_6)(\exists i_7)($	
$[\mathbf{Do}_{w_3 t_4},$	
$x_5,$	**družice** $\ldots (o\iota)_{\tau\omega}$
$[\mathbf{Perf}_{w_3}, x_6]$	**zaznamenat** $\ldots ((o(o\pi)(o\pi))_{\omega}\iota)$
$]$	**zrod** $\ldots (o\iota)_{\tau\omega}$
$\wedge\ x_5 \subset \mathbf{družice}_{w_3 t_4}$	**třetí** $\ldots \tau$
$\wedge\ [$	**Numerize** $\ldots ((o\iota)_{\tau\omega}(o\iota)_{\tau\omega}\tau)$
$[\mathbf{Of},$	**přechodně** $\ldots ((o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$
$[\mathbf{Numerize}, \mathbf{zrod}, \mathbf{třetí}],$	**trvající** $\ldots ((o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$
$[$	**radiační** $\ldots ((o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$
$[\mathbf{přechodně}, \mathbf{trvající}],$	**pás** $\ldots (o\iota)_{\tau\omega}$
$[\mathbf{Of},$	**země** $\ldots (o\iota)_{\tau\omega}$
$[\mathbf{radiační}, \mathbf{pás}],$	**Of** $\ldots ((o\iota)_{\tau\omega}(o\iota)_{\tau\omega}(o\iota)_{\tau\omega})$
$\mathbf{země}$	**družice** $\ldots (o\iota)_{\tau\omega}$
$]$	**Anytime** $\ldots (o\tau)$
$]$	**Onc** $\ldots ((o(o\tau))\pi)_{\omega}$
$]_{w_3 t_4},$	**Do** $\ldots (o(o\iota)(o(o\pi)))_{\tau\omega}$
i_7	**Perf** $\ldots ((o(o\pi))(o(o\pi)(o\pi)))_{\omega}$
$]$	**P** $\ldots ((o(o(o\tau))(o\tau))\tau)$
$\wedge\ x_6 = [\mathbf{zaznamenat}, i_7]_{w_3}$	(verbal object) $x_6 \ldots (o(o\pi)(o\pi))$
$)$	
$],$	
$\mathbf{Anytime}$	
$] \ldots \pi$	

Fig. 2. An example TIL logical analysis in the corpus of TIL constructions. The English equivalent of the analysed sentence is "*Satellites recorded the birth of the third temporarily-sustaining radiation belt of the Earth.*"

4 Conclusions and Future Directions

To conclude, we may say that the OTT as such provides a complex theoretical framework for the NLP research within which all relevant components (syntax, semantics, pragmatics) are present. Moreover, the OTT is open to the methodological variability in the present-day NLP since some of its parts rely on rule-based techniques (particularly TIL) while in other parts statistical methods and machine learning (parsing) can be used. In this respect, the OTT may be characterized as a hybrid approach.

We would like to stress one more point: thanks to the TIL formalism, the OTT can serve as a formal tool for handling knowledge representations and inference. In this respect, the OTT can be regarded as a basis for future knowledge-rich question answering systems based on full logic of the underlying discourse.

Acknowledgements. This work has been partly supported by the Czech Science Foundation under the project GA15-13277S.

References

1. Svoboda, A., Materna, P., Pala, K.: An ordered-triple theory of language. Brno Studies in English **12** (1976) 159–186
2. Svoboda, A., Materna, P., Pala, K.: The ordered-triple theory continued. Brno Studies in English **13** (1979) 119–165
3. Chomsky, N.: Aspects of the Theory of Syntax. Volume 11. MIT press (2014)
4. Sgall, P., Nebeský, L., Goralčíková, A., Hajičová, E.: A Functional Approach to Syntax: in Generative Description of Language. Elsevier, New York (1969)
5. Dowty, D.R.: Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ. Reidel (1979)
6. Montague, R.: Formal Philosophy. Yale University Press, New Haven (1974)
7. Morris, C.: Logical Positivism, Pragmatism and Scientific Empiricism. Number 449 in Exposés de philosophie scientifique. Hermann et Cie (1937)
8. Tichý, P.: The foundations of Frege's logic. Walter de Gruyter, New York (1988)
9. Palová, I.: The syntactic analyzer for Czech. In: Papers at the Conference on Cybernetics, Prague (1976)
10. Winograd, T.: Understanding natural language. Cognitive psychology **3**(1) (1972) 1–191
11. Šmerk, P.: Fast morphological analysis of Czech. In: RASLAN 2009, Recent Advances in Slavonic Natural Language Processing. (2009) 13–16
12. Jakubíček, M., Kovář, V., Šmerk, P.: Czech morphological tagset revisited. In: Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2011. (2011) 29–42
13. Cihanek, P.: The semantic analyzer for Czech. In: Papers at the Conference on Cybernetics, Prague (1976)
14. Horák, A., Pala, K.: Pavel Materna a třísložková teorie jazyka (in Czech, Pavel Materna and the Ordered-Triple Theory). Studia Philosophica **62**(2) (2015) 114–119
15. Chrz, T.: A system for knowledge representation based on intensional logic. Computers and Artificial Intelligence (3, 4) (1984) 193–209, 305–317
16. Horák, A.: The Normal Translation Algorithm in Transparent Intensional Logic for Czech. PhD thesis, Faculty of Informatics, Masaryk University, Brno (2002)
17. Kadlec, V.: Syntactic analysis of natural languages based on context-free grammar backbone. PhD thesis, Masaryk University, Faculty of Informatics (2008)
18. Horák, A.: Computer Processing of Czech Syntax and Semantics. Tribun EU (2008)
19. Jakubíček, M.: Extrakce strukturních informací z běžného textu na základě syntaktického analyzátoru (in Czech, Structural Information Extraction from Common Texts Based on Syntactic Parser). Master's thesis, Masaryk University, Faculty of Informatics (2008)
20. 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)
21. Kovář, V., Horák, A., Jakubíček, M.: How to analyze natural language with transparent intensional logic? RASLAN 2010, Recent Advances in Slavonic Natural Language Processing (2010) 69–76

22. Hlaváčková, D.: Databáze slovesných valenčních rámců VerbaLex (in Czech, VerbaLex – the Database of Verb Valency Frames). PhD thesis, Masaryk University, Faculty of Arts (2008)
23. Hlaváčková, D., Horák, A., Kadlec, V.: Exploitation of the verbalex verb valency lexicon in the syntactic analysis of czech. In: Proceedings of Text, Speech and Dialogue 2006, Brno, Czech Republic, Springer-Verlag (2006) 79–85
24. Horák, A., Pala, K., Duží, M., Materna, P.: Verb Valency Semantic Representation for Deep Linguistic Processing. In: Proceedings of the Workshop on Deep Linguistic Processing, ACL 2007, Prague, Czech Republic, the Association for Computational Linguistics (2007) 97–104