

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

Aleš Horák 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

RASLAN 2017
December 1, 2017

Outline

- 1 Ordered-Triple Theory
- 2 Semantics
- 3 Computer Tools
- 4 Continuation at FI MU
- 5 Conclusions

The Ordered-Triple Theory of Language (OTT)

- historical perspectives of the **Ordered-Triple Theory of Language** (OTT)
- authors: *Materna, Pala* and *Svoboda* at FF UJEP (1976-79)
- OTT captures **three fundamental components** of a language system: **syntax**, **semantics** and **pragmatics**
- fully comparable with similar linguistic theories (Chomsky, Sgall)
- systematic interconnection of **logic**, **linguistics** and **informatics**
- continuation of the mutual cooperation from 1995 at FI MU: Pala and Materna, later Horák
- the subsequent milestones related to OTT
- applying the **transparent intensional logic** (TIL)
- OTT relation to the natural language processing (primarily Czech)

Main motivation for developing the OTT

- reaction to the two influential language theories at this time (1976)
- to Chomsky's **generative grammar approach** (1964)
- to the **Prague functional generative framework (FGD)** by Sgall (1969)
- OTT had not been conceived explicitly as generative
- it allowed to deal with both **recognoscative and generative devices**
- Prague's FGD from the beginning contained **semantic component**
- in the form of the **tectogrammatical level** based on a set of **actants** (semantic roles)
- Chomsky's generative grammars were primarily based on **syntax**

The main features of the Ordered-triple Theory

- OTT offers a **theoretical framework** for a formal NL description
- provides all the **basic components** of any semiotic system (Morris)
- captures relations between **language user and real world**
- also relations between **language and language user** (pragmatics)
- NL expressions consist of syntactic, semantic and pragmatic constituents
- they can be described as **ordered triples** comprising:

(semantic component, formal language expression, pragmatic component)

Semantics – TIL

- the semantic part of OTT is conceived as strictly **intensionalistic**
- **problems** of the extensional approach to semantics are **explained**
- reasons for handling **referential phenomena** like individual roles, propositional attitudes, or episodic verbs are given
- the **logical analysis** of natural language in OTT relies on **TIL**
- proposed in 1970s by **P. Tichý**
- after August 1968 he left ČSSR, lived shortly in UK
- then moved to New Zealand to the Otago University in Dunedin (**died 1994**)

Semantics – Types

- the basic concepts of the intensional semantics consists of the **epistemic base**
- constituted by four sets: the **universe** (ι , set of individuals), the **set of truth values** (\circ), the **set of possible worlds** (ω) and the (continuous) **set of time moments** (τ).
- the simple type theory is used to produce derived entities
- the most typical intensions are given using the operation called **intensional descent** – 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).

Semantics – Constructions

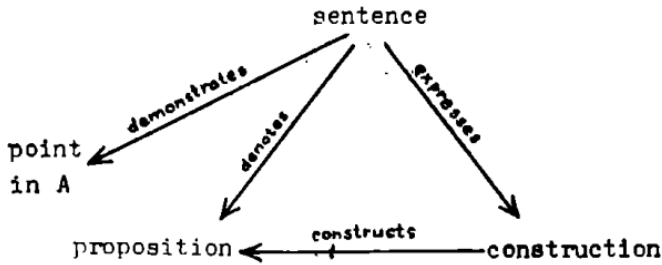
- further relevant concepts are: **constructions** (atoms, applications, abstractions), the relation between language expression, construction and intension
- class of **language constructions** is distinguished among all other constructions
- they can be expressed by **natural language expressions**
- 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
- grammar contains syntactic rules having the form of **context-free rules** plus semantic rules operating on the output of the grammar providing formulae of the λ -calculus as a result

Semantics – Time and Space

- the sets consisting of the universe, truth values and possible worlds are extended with further sets
- the **set of the time moments** allows to capture time characteristics including **grammatical tenses**
- the semantics of locational adverbs can be examined by adding the **set of space points**
- to handle **deictic (indexical) expressions** as e.g. personal pronouns
- to establish relations between **external pragmatics and semantics**.

Semantics – Pragmatics

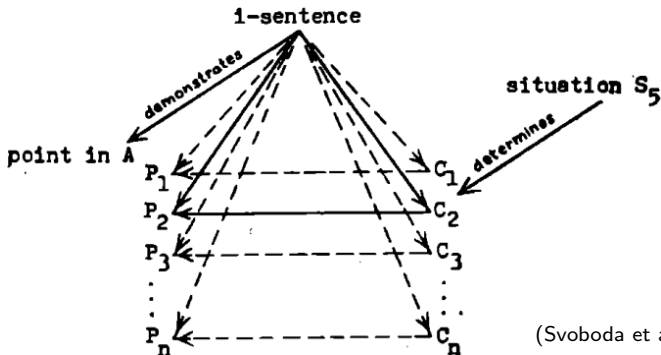
- attention is paid to the semantic relations of **expressing**, **denoting** and **constructing**
- to the pragmatic relations of **demonstrating** (internal pragmatics) and **determining** (external pragmatics)



(Svoboda et al, 1979)

Semantics – Pragmatics

- attention is paid to the semantic relations of **expressing**, **denoting** and **constructing**
- to the pragmatic relations of **demonstrating** (internal pragmatics) and **determining** (external pragmatics)

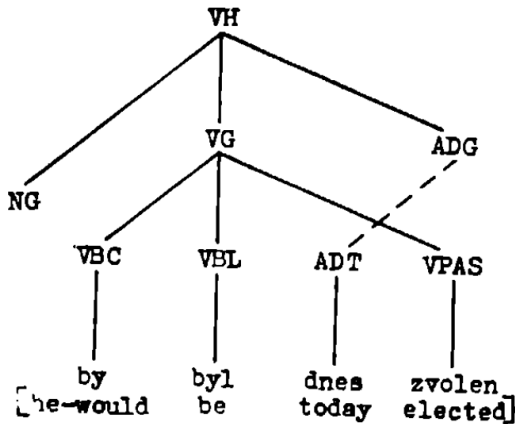


Computer Tools for OTT

- the first version of OTT contained a context-free like grammar
- a **set of procedural rules for Czech** (Palová 1976)
- based on the **procedural grammar** inspired by T. Winograd for English (**SHRDLU (1972)**)
- it was written in **LISP 1.5**, contained 34 LISP functions
- tested on **mainframe TESLA 200** at the **VUT Computer Lab**

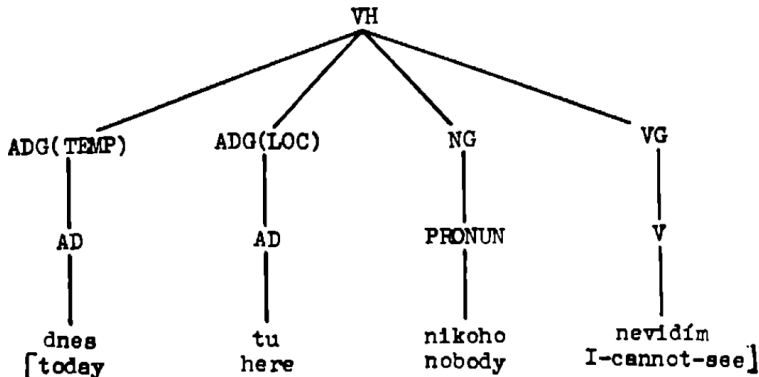


Procedural Grammar



(Svoboda et al, 1979)

Procedural Grammar



(I can't see anybody here today)

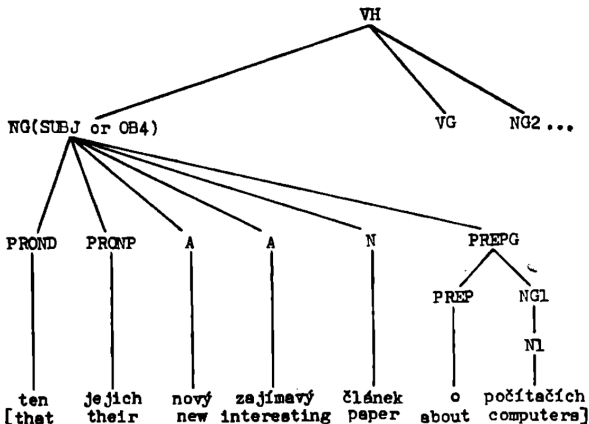
(Svoboda et al, 1979)

Morphology

- in 1976 no standard morphological analyzer for Czech existed
- a **morphological input** for the Czech procedural grammar was prepared **manually**
- it had a form of the “syntactic dictionary” with **corresponding parts of speech and lists of grammatical features**
- this file was **almost identical** with the output of the present-day morphological analyser **like Majka**

Syntactic Analyser

- The procedural grammar produced a syntactic structure of a sentence in the form of the **labelled tree graph**

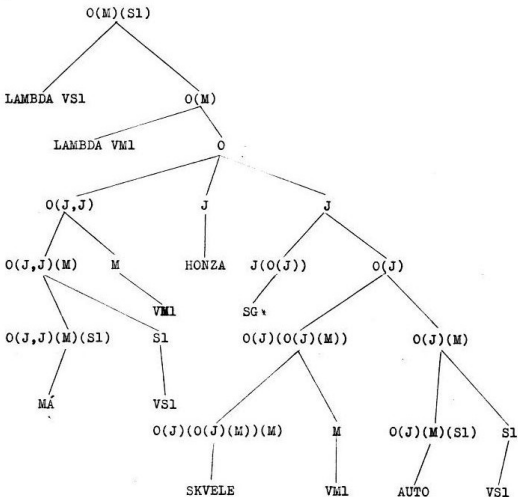


(Svoboda et al, 1979)

Semantic Analyser

- trees served as an **input for the semantic analyser** (Čihánek, 1976)
- also for rules handling **attitudes** of language users in the internal pragmatics framework
- both is **missing** in Chomsky's and Prague approaches
- semantic analyser was implemented in **LISP 1.5** and tested on TESLA 200 as well.
- The output was a tree structure **representing a λ -calculus formulae** obtained from the syntactic tree of the Czech sentence
- λ -calculus formulae are corresponding to the **natural language constructions**

Semantic Tree Structure



(Čihánek, 1976)

Historical remark

- A historical remark: described results were achieved **until 1989**
- in this year A. Svoboda **had to move** to Opava University
- Materna and Pala **started to teach** at the FI MU
- the work on OTT thus **continued** there
- in **1997** the **Laboratory of NLP** was established at the FI MU

Historical remark

- A historical remark: described results were achieved **until 1989**
- in this year A. Svoboda **had to move** to Opava University
- Materna and Pala **started to teach** at the FI MU
- the work on OTT thus **continued** there
- in **1997** the **Laboratory of NLP** was established at the FI MU
- This means that we can celebrate **20 years** of the NLP Centre at the Faculty of Informatics MU



Historical remark

- The Faculty of Informatics MU was established in 1994
- **Materna** and **Pala** started teaching there in 1995
- Materna offered **lectures about TIL**
- attracted students interested in logical analysis of natural language, e.g. **A. Horák**
- he dealt with the problems of the **logical and semantic analysis** of language in his *diploma thesis* and *dissertation*
- he made a progress in TIL, in an important part of OTT and continued with the computer implementation
- enabling to translate common natural language sentences into the **TIL constructions**

Continuation of OTT at FI MU

- Horák's research brought a new original result – the **Normal Translation Algorithm** (2002)
- plus the **syntactic analyser Synt** for Czech (2008)
- context actions fo **translating syntactic trees** to corresponding intensional constructions
- **further cooperation** with P. Materna and M. Duží, GAČR 2005–2007, 2010–2012, 2015–2017 (FI MU + VŠB-TU Ostrava)
- **large corpus of TIL logical constructions** (2010)
- the corpus is presently suitable for explicating various language phenomena in common Czech texts
- more than **5,000 sentences** semi-automatically analysed and translated according to NTA

Corpus of TIL Constructions

$$\begin{aligned}
 & \lambda w_1 \lambda t_2 [P_{t_2}, \\
 & \quad [Onc_{w_1}, \\
 & \quad \quad \lambda w_3 \lambda t_4 (\exists x_5)(\exists x_6)(\exists i_7) (\\
 & \quad \quad \quad [Do_{w_3 t_4}, \\
 & \quad \quad \quad \quad x_5, \\
 & \quad \quad \quad \quad [Perf_{w_3, x_6} \\
 & \quad \quad \quad] \\
 & \quad \quad \quad \wedge x_5 \subset \text{družice}_{w_3 t_4} \\
 & \quad \quad \quad \wedge [\\
 & \quad \quad \quad \quad [Of, \\
 & \quad \quad \quad \quad \quad [Numerize, \text{zrod, třetí}], \\
 & \quad \quad \quad \quad \quad [\\
 & \quad \quad \quad \quad \quad \quad [přechodně, trvající], \\
 & \quad \quad \quad \quad \quad \quad [Of, \\
 & \quad \quad \quad \quad \quad \quad \quad [radiační, pás], \\
 & \quad \quad \quad \quad \quad \quad \quad \quad \text{země} \\
 & \quad \quad \quad \quad \quad \quad] \\
 & \quad \quad \quad \quad \quad] \\
 & \quad \quad \quad]_{w_3 t_4}, \\
 & \quad \quad \quad i_7 \\
 & \quad \quad] \\
 & \quad \quad \wedge x_6 = [\text{zaznamenat}, i_7]_{w_3} \\
 & \quad) \\
 &], \\
 & \text{Anytime} \\
 &] \dots \pi
 \end{aligned}$$

Družice zaznamenaly zrod třetího přechodně trvajícího radiačního pásu Země.

Satellites recorded the birth of the third temporarily-sustaining radiation belt of the Earth.

družice ... $(ol)_{\tau\omega}$
zaznamenat ... $((o(o\pi))(o\pi))_{\omega\iota}$
zrod ... $(ol)_{\tau\omega}$
třetí ... τ
Numerize ... $((ol)_{\tau\omega}(ol)_{\tau\omega}\tau)$
přechodně ... $((ol)_{\tau\omega}(ol)_{\tau\omega})$
trvající ... $((ol)_{\tau\omega}(ol)_{\tau\omega})$
radiační ... $((ol)_{\tau\omega}(ol)_{\tau\omega})$
pás ... $(ol)_{\tau\omega}$
země ... $(ol)_{\tau\omega}$
Of ... $((ol)_{\tau\omega}(ol)_{\tau\omega}(ol)_{\tau\omega})$
družice ... $(ol)_{\tau\omega}$
Anytime ... $(o\tau)$
Onc ... $((o(o\tau))\pi)_{\omega}$
Do ... $(o(ol)(o(o\pi)))_{\tau\omega}$
Perf ... $((o(o\pi))(o(o\pi)(o\pi)))_{\omega}$
P ... $((o(o\tau))(o\tau))_{\tau}$
 (verbal object) x_6 ... $(o(o\pi)(o\pi))$

Valency Frames and OTT

- One recent result in the NLP Centre is a **valency database for Czech language VerbaLex** (2006, 2008)
- It contains **approx. 10,500 Czech verbs** and it is the largest valency database for Czech
- verb valency frames represent verbs as **predicates with their arguments**
- thus they can be **linked with the language constructions in TIL** (2007)
- therefore VerbaLex can serve for building a list of the Czech verb predicates equipped with respective **logical types**
- in this respect it can and will be exploited in **further research**

VerbaLex

dokumentovat₂^{biasp} **zaznamenat**₃^{pf} **zachytit**₂^{pf}
zaznamenávat₃^{impf} **zachycovat**₂^{impf}

1 dokumentovat₂ ≈

2 dokumentovat₂, zachytit₂, zachycovat₂, zaznamenat₃, zaznamenávat₃ ≈

-frame: **AG** <person:1>_{obl} **VERB** **STATE** <state:4>_{obl} | **ABS** <abstraction:1>_{obl}
a1 i4

ART <creation:2>_{opt}
i7

-example: filmem dokumentoval život Inuitů (**biasp**)

-example: fotografií zaznamenal bídu Afriky (**pf**)

-example: hudbou zachytil folklór Hanácka (**pf**)

Conclusions

- OTT as such provides a **complex theoretical framework** for the NLP research
- provides all **relevant components** (syntax, semantics, pragmatics)
- OTT is open to the **methodological variability** in the present-day NLP
- some of its parts rely **on rule-based techniques** (particularly TIL)
- in other parts **statistical methods and machine learning techniques** (parsing) can be used
- in this view, OTT may be characterized as a **hybrid approach**

Possible applications of the OTT

- with TIL the OTT formalism can serve as a formal tool for handling **knowledge representations and inference**
- OTT can be regarded as a **basis for future knowledge-rich question answering systems** based on full logic of the underlying discourse (2017)
- see also the **experiments with TIL and AI agents** (Duží et al, Logic & Artificial Intelligence for Multi-Agent Systems, ongoing research)
- finally, we ask members of the NLP Centre to cite the OTT framework to make our results more **visible**