

Počítačové zpracování přirozeného jazyka a Transparentní intenzionální logika

Jiří Raclavský

Katedra filozofie Filozofické fakulty Masarykovy univerzity v Brně
e-mail: raclavsk@phil.muni.cz

Abstrakt

V této stati bych chtěl stručně pojednat o tématu zmíněném v názvu. Nejprve bude charakterizována Transparentní intenzionální logika a dílo Pavla Tichého, poté bude vytčeno několik základních cílů, které od počítačového zpracování očekáváme, přičemž se budeme zabírat současným stavem vypracování těchto cílů v rámci užívaným Transparentní intenzionální logiku.*

1 Logický aparát – Transparentní intenzionální logika (TIL)

TIL je životní dílo českého logika Pavla Tichého (*1936 Brno, †1994 Dunedin), který od konce šedesátých let působil v zahraničí, zejména na Novém Zélandu. Základ TIL byl budován již v druhé polovině 60. let dvacátého století, první publikovaná verze je z roku 1971, takže přišla vzhledem ke konkurenční intenzionální logice R. Montagua jen s drobným zpožděním, a dlužno říci, že se jí mezinárodně stále nedostává adekvátního ohlasu. Jak Tichý, tak i Montague adaptovali Churchův lambda kalkul, nicméně nestejným způsobem. Montague se přidržel kontextualismu (v návaznosti na Fregeho a Carnapa), podle kterého výraz označuje extenzi či intenzi v závislosti na tom, v jakém se nachází kontextu. Toto mezinárodně přijímané paradigma Tichý nepřijal a výrazy dle něj

* Na konferenci *Organon V.* byla v rámci stejnojmenného příspěvku podávána též informace o tzv. editoru konstrukcí, což je internetové rozhraní pro sestavení formulí TIL (a jejich parsování) a jejich ukládání do nějakého datového úložiště.

označují, denotují stále týž druh entity (*antikontextualismus*), tj. výrazy empirické označují intenze, výrazy neempirické extenze. Další problémy s tím přímo souvisí: je potřeba nějak formálně vystihnout, jak a kdy např. „Jitřenka“ denotuje Venuši. Tichý přesvědčivě zdůvodnil, že operátory Montaguovy (a jeho následovníků) jsou chybně definované, takže následně se tyto intenzionální logiky neslučují s principem extenzionality či principem kompozicionality. Fundamentální rysem Tichého logiky je modelování intenzi pomocí λ -termů, které determinují funkce z možných světů, tj. intenze, např. $\lambda w[...w...]$ (*explicitní intenzionalizace*). Rozdílnost signifikace výrazů jako „Jitřenka“ má exaktní formální pendant v tom, zda příslušný λ -term intenze je, či není aplikován na argument možného světa: pokud ano, výraz „Jitřenka“, který sice stále denotuje jistou intenzi, navíc tu slouží k referenci na hodnotu této intenze v daném světě (Venuši, či, v jiném světě, Mars). Tato pečlivá práce se supozicemi de dicto/de re řeší ty největší potíže, s nimiž se všechny jiné intenzionální logiky (včetně modálních) dodnes neúspěšně potýkají. TIL je dvouhodnotová logika, ale pracuje s parciálními funkcemi.

Dalším významným rysem TIL je důsledně objektuální pojetí (*anti-formalismus*), které nemá daleko k původnímu Churchově označovacímu schématu:

výraz	
vyjadřuje	
smysl	Tichý: konstrukci
determinuje (výraz denotuje)	
denotát	

Pro případ empirických výrazů – jsou-li tyto v supozici de re – tu máme ještě ono *referování* na hodnotu intenze v daném světě. Koncem 70. let přidal k modálnímu parametru ještě parametr temporální, takže TIL může být chápán jako intenzionální logika s temporálním parametrem. Také může být TIL chápán jako objektuálně pojatý λ -kalkul s typy (atomickými typy jsou individua, pravdivostní hodnoty, možné světy, časové okamžiky/reálná čísla). Vrátime-li se k objektualitě, λ -termy nejsou vnímány jen jako syntaktické pomůcky, jsou to zápisy konstrukcí daných entit. Kategorie konstrukcí byla Tichým vyzdvihována až od poloviny 80. let, ovšem byla přítomna již v raných verzích TIL. Konstrukce jsou, módně řečeno, strukturované významy, procedury, hyperintenze. Konstanty λ -kalkulu pojal Tichý zcela objektuálně jakožto triviální procedury-konstrukce zvané trivializace (tvar 0X), proměnné také nepovažoval za pouhá písmena, ale za svébytné procedury-konstrukce, aplikaci funkce na argument vyložil jako tzv. kompozici ($[F A_1 \dots A_n]$),

λ -abstrakci jakožto λ -uzávěr. Významy vět (věty denotují propozice) pak typicky jsou konstrukce druhu uzávěru ($\lambda w[\lambda t[\dots w \dots t \dots]]$), které konstruuji funkce z možných světů a časů do pravdivostních hodnot atd. Ona zvláštní konstrukce druhu trivializace umožňuje vystihnout *zmiňování* konstrukcí („pojmu“), srov. třeba užití pojmu kůň ve větě „Hatátitla je kůň“, a zmínění tohoto pojmu ve větě „Kůň je pojem“. Ve vrcholné koncepci TIL z konce 80. let Tichý dospěl k adaptaci *rozvětvené teorie typů* (která brání bludnému kruhu), takže můžeme TIL charakterizovat jako „higher-order intensional logic“. Dá se také říci, že Tichému se podařilo skloubit a domyslet klíčové ideje Fregeho, Russella, Churcha. Pro bližší informace viz zejm. [6].

2 Tichého dílo

Dovolím si zde učinit pokus o rozčlenění Tichého díla, které čítá vlastně tři knihy a 46 statí:

(IIL) – *Introduction to Intensional Logic* (1976), nepublikována,

(FFL) – *Foundations of Frege's Logic* (1988),

(MDG) – „*Meaning-Driven Grammar*“, nedokončena, publikován úvod.

Statě jsou přetištěny v *Pavel Tichý's Collected Papers in Logic and Philosophy*, 2004. Tichý viděl největší oblast aplikability TIL v oblasti *logické analýzy přirozeného jazyka* (LANL), zvláště MDG měla být analýzou podstatného fragmentu angličtiny, s tím, že Tichý při svém očekávaném návratu předpokládal počítačové zpracování. Kromě LANL však TIL prokazuje značnou schopnost i v oblasti analytické filozofie, která souvisí s jazykem, dokonce tu můžeme hovořit o ambici explikovat „celý pojmový rámec“. Nyní tedy ony okruhy:

LANL čistě – MDG, Questions, Answers and Logic (1978), The Logic of Temporal Discourse (1980), The Semantics of Episodic Verbs (1980);

LANL + zdůvodnění (filozofie jazyka) – IIL, FFL, An Approach to Intensional Analysis (1971), What Do We Talk About? (1975), Two Kinds of Intensional Logic (1978), De dicto and de re (1978), The Transiency of Truth (1980), Constructions (1986), 3× o subjunktivních kondicionálech, Cracking the Natural Language Code (1994), The Analysis of Natural Language (1994), aj.;

dedukce (pouze pro jednoduchou teorii typů)¹ – ILL, The Foundations of Partial Type Theory (1982), The Logic of Ability, Freedom, and Responsibility (1982), Indiscernibility of Identicals (1986);

logika (mimo dedukci) – Smysl a procedura (1968), Intensions in Terms of Turing Machines (1969),² On the Vicious Circle in Definitions (1971), Eine Exposition des Gödelschen Unvollständigkeitsbeweises in der einfachen Typentheorie (1962), aj.;

filozofie – FFL, 3× verisimilitude, Einzeldinge als Amtsinhaber [Particulars as Things-to-be] (1987), Constructions as the Subject Matter of Mathematic (1995), Ability and Freedom (1983), ontologický důkaz, 2× kontra Frege, 1× kontra Plantinga, 2× kontra Kripke, 1× kontra Putnam, 1× kontra Wittgenstein, aj.

3 Současné okruhy aplikace a rozvoje TIL

Toto téma podáme jen velmi stručně a poněkud nepřesně (zvláště je třeba říci, že s TIL pracovalo v minulosti nemálo dalších lidí). Ze seznamu je patrné, že TIL má ohlas zejména v „Československu“. Z charakteristiky TIL, kterou jsme si uváděli výše, a ze srovnání s mezinárodními trendy je patrný i důvod ignorance TIL v mezinárodním prostředí – TIL se neslučuje s přijímanými paradigmaty. Znalci též říkají, že Tichý s TIL v mnohém předběhl dobu, takže se dá snad domnívat, že TIL se ohlasu dočká zpětně (třeba tak, jako se to stalo Fregeho dílu).

LANL + logika – Materna, Duží, Jespersen,

LANL počítačově – Horák (NTA), dílčí spolupracovníci,

multiagentní systémy – Duží a spolupracovníci (LabIS),

filozofie podkládající TIL, aplikace TIL do filozofie – Cmorej, Materna, Jespersen, Gahér, Kuchyňka, Raclavský,

databázové modelování (HIT), apod. – řada lidí,

dedukce – Müller.

¹ Jde v podstatě o gentzenovskou přirozenou dedukci.

² Právě v těchto dvou statích je uvedena myšlenka, že významy výrazů jsou strukturované pojmy, procedury.

Dodejme, že v současnosti dokončují Materna, Duží, Jespersen anglicky psanou knihu, která má být systematickou expozicí TIL. O Laboratoři inteligentních systémů (LabIS) zde nebudeme podávat jiné informace než prostý výčet (přejatý z homepage LabIS):

- multiagentní systémy;
- logická analýza přirozeného jazyka;
- specifikace jazyků založených na přirozeném jazyce;
- reprezentace znalostí;
- inferenční stroje non-monotonic reasoning a belief revision;
- process management, control a coordination;
- simulace procesů a predikce.

Níže se omezíme jen na problematiku implementace NTA, Normálního translačního algoritmu.

4 Co chceme od počítačového zpracování přirozeného jazyka

Domnívám se, že je přirozené po počítačovém zpracování přirozeného jazyka požadovat přinejmenším dosažení tři následujících cílů, které pracovníčně nazvu „sémantický stroj“, „translační stroj“, „inferenční stroj“.

4.1 Sémantický stroj

Úkolem sémantického stroje je mechanicky provést logickou analýzu výrazu přirozeného jazyka, tedy přiřazení analýzy, modelu významu synchronně daného jazyka, kterým je živý přirozený jazyk (synchronně daný jazyk modelován jako zobrazení z výrazů do významů). Cílovým „jazykem“ je umělý jazyk, tedy formalismus TIL. Poněkud lidově se explikaci přirozeného jazyka pomocí formálního jazyka říká „překlad“.

vstup:	výraz přirozeného jazyka (NLE)
uvnitř sém. stroje:	desambiguace, ukládání do databáze apod.
výstup:	formule logického aparátu (výrazů formálního jazyka)

Při dosahování tohoto cíle je TIL nejdále. Jádro teorie podal vlastně již Tichý, a ačkoli přirozený jazyk skýtá ještě mnoho oříšků, které jsou diskutovány v odborné literatuře, to zásadní je hotovo. Podstatný byl vznik „počítačové implementace“ této teorie, a právě tím je projekt „Algoritmus normální translace v TIL pro češtinu“, který začal běžet před rokem 2002 na Fakultě informatiky MU pod vedením Aleše Horáka. Co se týče praktických výsledků, dle informací je sémantický stroj TIL s to zpracovat přes 90 % korpusu češtiny. Ještě dodejme, že není známo, že by nějaký jiný logický aparát (aspirující provádět logickou analýzu přirozeného jazyka) než TIL disponoval sémantickým strojem.

4.2 Inferenční stroj

Co je úkolem inferenčního stroje je jistě zřejmé, jde tu o počítačově zpracovanou dedukci, přičemž je třeba odlišovat dedukci na úrovni významů (konstrukcí) a na úrovni výrazů. Některé obecné požadavky:

- důkaz vyplývání daného závěru z daných předpokladů
- odvození závěru (závěrů) k daným předpokladům
- nalezení důkazu (důkazů) k danému tvrzení

Pokud je mi známo, pro TIL je již v zásadě hotova teorie, na praktické části se pracuje v okruhu LabIS, lze proto říci, že inferenční stroj pro TIL „už už bude“, byť lze tušit, s jak obrovským objemem práce se nakonec vybudování inferenčního stroje bude muset setkat.

5 Translační stroj

Je zřejmé, že důležitým cílem, který má jasné komerční uplatnění, je automatický překlad mezi dvěma přirozenými jazyky. Základní idejí je, že dva výrazy různých jazyků mohou kódovat/vyjadřovat též význam (který nazýváme logickou analýzou). Schématicky:

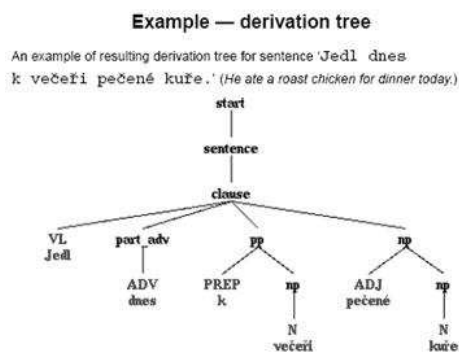
- vstup: výraz jednoho přirozeného jazyka (NLE1)
- uvnitř: a) sémantický stroj zpracuje NLE1
b) vyhledávač prohledá databázi a k analýze nalezne NLE2
- výstup: výraz druhého přirozeného jazyka (NLE2)

Pro případ TIL tu zatím není sémantický stroj analyzující např. angličtinu, je však zjevné, že tato záležitost není v budoucnu principiální překážkou.

6 Algoritmus normální translace (NTA) v TIL pro češtinu

Jak již bylo řečeno, sémantickým strojem pro TIL je Algoritmus normální translace (NTA) v TIL pro češtinu, který je implementován na brněnské Fakultě informatiky Masarykovy univerzity a je v podstatné míře dílem Aleše Horáka, pracujícího v Laboratoři zpracování přirozeného jazyka FI MU (vedené doc. Karlem Palou), kde mj. vyučuje i prof. Pavel Materna kurz „Logická analýza přirozeného jazyka“. Daný projekt byl součástí Horákovy disertační práce „The Normal Translation Algorithm in Transparent Intensional Logic for Czech“, jež byla dokončena v roce 2002. Projekt, na kterém jsou zaměstnání i další spolupracovníci, je v současnosti podporován grantem GAAV „Překlad českých vět na konstrukce transparentní intenzionální logiky“.

V základě stojí syntaktický analyzátor, který zpracovává vstupy, získané výrazy češtiny, a podrobuje je „lingvistické analýze“. Dle popisu je to head-corner chart parsing (tabulkový analyzátor s řídicím prvkem) s přidáním kontextovými akcemi a testy pro zajištění gramatické shody a generování výstupu ve tvaru stromu závislostí (general lexical tree adjoining grammar parser). Pro češtinu je důležité zohlednění teorie slovesné valence a je činěna kooperace s Prague Dependency Tree-Bank. Z disertace [2] přejímáme pro ilustraci následující obrázek:



Obrázek 1.

Významným rysem Horákovy disertace je rozřídění analýz TIL nacházejících se v různých odborných publikacích podle klasifikace slovních druhů. Velice důležité je zapracování obsahu klíčových Tichého statí týkajících se LANL, tj. Tichého analýzy slovesných časů, vidů apod. K nim

Horák přidal i nové analýzy, protože ne všechny slovní druhy byly dosud v odborných pracích TIL pokryty, jedná se zejména o adverbial modification, noun prepositions, pomocná a modální slovesa. Samozřejmostí je parsování formulí (závorky apod.), také parsování typů. Pro ilustraci ještě obrázek:

Example — logical analysis

```
evaluation of rule_schema for np 'pečené kuře'
4, 6, -npnl -> .{ left_modif } np .: klgNnSc145
agree_case_number_gender_and_propagate OK
rule_schema: 2 nterms, 'lwtx(awtx(#1) and awtx(#2))'
And constra, Abstr and Exi vars are just gathered
1 (1x1) constructions:
 $\lambda w_2 \lambda t_3 \lambda x_4 ([\text{pečený}_{w_2 t_3}, x_4] \wedge [\text{kuře}_{w_2 t_3}, x_4]) \dots (\alpha)_{\tau \omega}$ 
And constra: none added
Exi vars: none added
```

Obrázek 2.

Výsledky analýz jsou ukládány do databáze, „tiskové“ výstupy jsou možné ve formátech HTML, \LaTeX , PostScript, PDF. Pro další informace odkazujeme na práci [2], která je dostupná skrze internetový protokol http.

Reference

- [1] Horák, A., *Algoritmus normální translace (NTA)*.
<http://www.fi.muni.cz/~hales/disert/>
- [2] Horák, A., *The Normal Translational Algorithm in Transparent Intensional Logic for Czech*, disertační práce, Fakulta informatiky Masarykovy univerzity v Brně, Brno 2002.
- [3] *Laboratoř inteligentních systémů (LabIS)*.
<http://labis.vsb.cz/>
- [4] *Laboratoř zpracování přirozeného jazyka FI MU*.
<http://nlp.fi.muni.cz/>
- [5] Tichý, P., *Introduction to Intensional Logic*, 1976 (nepublikovaný manuskript).
- [6] Tichý, P., *The Foundations of Frege's Logic*, Walter de Gruyter, Berlin–New York 1988.
- [7] Tichý, P., *Pavel Tichý's Collected Papers in Logic and Philosophy*, in V. Svoboda & B. Jespersen & C. Cheyenne (eds.), Otago UP–Filosofia, Dunedin–Praha 2004.
- [8] *Transparentní intenzionální logiky (TIL)*, homepage.
<http://til.phil.muni.cz/>