

## **Problémy komunikace s počítačem v přirozeném jazyce**

Předloženou práci lze chápat jako jakýsi povšechný úvod do obecně a velmi široce pojaté tematiky obsažené v názvu práce. Autorka svou práci člení na dvě větší části: a) o jazyce obecně a o jazycích přirozených a umělých; b) o aspektech umělé inteligence spjatých s přirozeným jazykem, tj. o reprezentaci znalostí, některých základních typech logik, sémantických sítích a produkčních systémech. Cílem práce je ukázat, jak obtížné, ba nemožné je za současného stavu poznání přimět počítač alespoň vzdáleně se přiblížit tomu, jak jazyku rozumějí a jak jazyk produkují lidé. Autorka ve své práci shrnuje základní, všeobecně známé problémy automatické analýzy a syntézy přirozeného jazyka. Tyto problémy stručně a ne vždy úplně přesně charakterizuje. Její práce je povrchním kompilátem notoricky známých, starých a víceméně elementárních faktů v oboru (matematická) lingvistika a umělá inteligence a nepřináší nic nového, ovšem v diplomové práci se – na rozdíl třeba od práce disertační – nic objevného obecně nepředpokládá. Lze říci, že autorka s pojmy pracuje někdy nepřesně, nepromýšlí je do hloubky. Z práce pro mne vyplývá, že zná jen některé elementární úvodní texty v oboru lingvistika a matematická lingvistika, staré pár desítek let a že skoro vůbec nereflektuje moderní vývoj přinejmenším v (matematické) lingvistice. O řadě problémů v práci pojednaných se ve světle moderního jazykovědného výzkumu a informatiky a jistě i umělé inteligence dá přece psát mnohem zajímavěji, než předvedla autorka. Možná by bylo bývalo vhodnější, kdyby si vybrala téma méně široké a ponořila se do nějakého konkrétního problému.

Podívejme se nyní podrobněji na jednotlivé části práce, s nimiž jsem měl problémy.

### **2. Přirozený jazyk**

s. 2, poslední odstavec: ... dnes existuje *nepřeberné* množství jazyků ... Co znamená *nepřeberné*? Ví autorka aspoň přibližně, kolik jich je, i u vědomí nejasných hranic mezi jazykem (centrálním dialektem) a dialekty? Jsou to tisíce, desítky či stovky tisíc?

### **2. Struktura jazyka**

Není pravda, že *systém jazyka se dá rozložit na syntax, sémantiku a pragmatiku*. Co fonetika, fonologie, morfologie (např. právě v češtině, ve většině slovanských jazyků, latině, řečtině atd. velmi složitá), stylistika, textová lingvistika, problematika diskursu? Přece fakt, že čeština má pět slovesných konjugačních tříd, je součástí češtiny jako systému. Například 3. osoba singuláru přítomného končí u českých sloves jen na *-e, -ě, -í, -á, jest*; tento systémový fakt se týká výhradně *flektivní morfologie*.

Opravdu se platí, že: ... *stejně jako jedinečným výkladem významů, se zabývá pragmatika*? Co je to jedinečný výklad významů? Jak by autorka třeba chápala větu: *Nemohla bys, Šárko, už konečně zavřít okno? Co je to tu za čurbes?* Dal by se na tomto příkladě osvětlit pojem *pragmatika*?

Kam by se do jazykového systému mělo zařadit aktuální členění zničehonic uvedené na konci s. 3?

V práci jsou některé notoricky známé příklady různých jazykových jevů. Autorka si je jistě nevymyslela, nebylo by tedy vhodné uvést, odkud přesně jsou a nejen uvádět obecně použitou literaturu?

s. 25: popis *metonymie* je nepřesný. Jak to autorka myslí? Pozor na příklady!

### 3.4.4 Vágnost

*vágnost ... fenoménem..., který je však v teoretické oblasti zpravidla považován za nežádoucí. Co je to teoretická oblast?*

ad s. 26, 1. řádek: vágnost je nepravidelnost? V jakém smyslu?

ve 2. odstavci: *Např. přítomný čas slovesa zahrnuje vždy současnost.* V jakém smyslu? Co když máme třeba věty:

*Nevěděl, co dělám. Nikdy nevěděl, co mám za lubem.  
Nebude vědět, co dělám.*

Zvažte dále třeba *prézens historický*. V jakém smyslu tam jde o současnost?

### 3.6 Neúplná informace

na konci: *srovnávání na nadřazenost, srovnávání na nadčást.* Zvláštní vazba *na* a zvláštní slovo *nadčást*.

## 4. Umělé jazyky

Ve třetím odstavci se objevuje slovo *řeč* vedle slova *jazyk*. Jaký má slovo *řeč* význam? Jak se má k cizím pojmům *langue, parole, langage*?

Nevhodně mnoho místa se věnuje Orwellovu newspeaku, vedle jediného odstavce věnovanému esperantu to je nevyvážené. Angsocem nikdo nikdy nemluvil, esperanto je užívaný jazyk.

### 4.1 Teoretické předpoklady

Předpoklady čeho? Vytvoření umělého jazyka? Pokud ano, má se to explicitně uvést v názvu!

### 4.2 Srovnání s přirozeným jazykem

na konci s. 35: opět směšování pojmu *řeč* a pojmu *jazyk*. Autorka jich nesprávně užívá jako synonym.

## 5. Reprezentace znalostí

poslední věta na s. 40: V jakém smyslu *Na aktuálním členění závisí i určení, k jakému antecedentu se vztahuje odkazovací zájmeno?* Vysvětlete, prosím, a uveďte příklady.

Prosím, aby se otázky, jež výše vznáším, probraly při obhajobě.

Po stylistické a zejména pravopisné stránce je práce psána dost nedbale, mám dojem velkého spěchu, s jakým autorka svou kompilaci „spíchla“. Občas bývá špatně interpunkce, občas překlep –

na s. 4: ... *volný slovosled je volný v tom smyslu, že není vázán rozložením větných členů*. Mohla by autorka, prosím, vysvětlit, co to znamená?

### 2.3 Porozumění jazyku

ad úvaha o jazyku a mozku: *jeho [jazyka] užívání probíhá na základě nevědomých procesů*. Jak tuto větu autorka chápe? Prosím vysvětlit.

na s. 9: *Dalším problémem porozumění je skladba ...* Co je zde *skladba* a jak se má k pojmu *syntax*? Jaké jevy ilustrují příklady na s. 9?

### 2.4 Analýza textu

Autorka omezuje systémy pro zpracování přirozeného jazyka na převod *povrchového tvaru jeho výrazů do některého druhu reprezentace* či ... *zpracovat vstupní text v přirozeném jazyce a přiřadit mu určitou strukturu ve výstupním jazyce*. Co takhle například značkování jazykových korpusů? Nebo jen vytvoření korpusu?

### 3. Problémy s porozuměním přirozenému jazyku

na konci s. 15: *Chybný, znalostem světa odporující význam by bylo možné vyloučit zavedením podrobné soustavy sémantických rysů podstatných jmen, i když by se tím vlastní lingvistická analýza překročila*.

Ptám se, proč jenom podstatných jmen a proč by se tím vlastní lingvistická analýza překročila?

### 3.4 Sémantické problémy

s. 21 na konci 2. odstavce: *jeden pojem má více významů... jeden význam se dá říci více pojmy* — užil bych tu snad spíše *slovo* než *pojem*? Nejde přitom jen o homonymii a synonymii slov, ale i celých konstrukcí. Uměla by autorka uvést příklady strukturní homonymie a synonymie (v jejím textu jsou ostatně některé uvedeny) a problém homonymie a synonymie trochu rozvést z hlediska forem a funkcí a asymetrického dualismu?

#### 3.4.1 Synonymie

První věta zní: *Synonyma jsou slova nebo slovní spojení se vzájemně stejným nebo podobným významem...* Co je to *vzájemně stejný* – trochu zvláštní formulace!?

V odst. 4 se najednou operuje specifickou terminologií pražského Funkčního generativního popisu (*tektogramatická struktura, funktor, gramatém*), neškodilo by tyto pojmy nějak předem vysvětlit! Co vlastně znamenají?

#### 3.4.2 Homonymie

Definice je nepřesná, proč se v definici homonyma operuje *původem* slova? Jak to autorka myslí? Mohla by autorka uvést vyčerpávajícím způsobem typy homonymie? Zná speciálnější pojmy *homofonie* a *homografie*?

Na počátku 3. odstavce: ... *vyřazení homonym ze slovníku*? Co to znamená?

tyto nedostatky by se daly při větší pečlivosti u tak krátké práce odstranit. Jen namátkou uvádím některé nedostatky:

s. 2, uprostřed: ... *jedinec ve svém životě získá z jeho vlastní interakce* ...  
správně má být: *jedinec ve svém životě získá ze (své) vlastní interakce* ...

s. 3, 3. řádek 2. odstavce: *Kontext však tuto osobu zmiňuje a proto nebylo* ... [chybí čárka]  
správně má být: *Kontext však tuto osobu zmiňuje, a proto nebylo* ...

s. 4, 13. řádek odspodu: *Informace je objektivní zatímco znalost* [chybí čárka]  
správně má být: *Informace je objektivní, zatímco znalost*

s. 6, 10. řádek odspodu: *Heuriskein*, správně má být: *heuriskein*, není to přece řecké vlastní jméno

s. 13, poslední řádek celého většího prvního odstavce: *různý výrazy*, správně má být *různé výrazy*

s. 14, 4. řádek shora: *označuj*, správně má být *označuje*

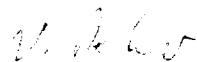
s. 16, 6. řádek odspodu: *Tam, kde je to zapotřebí vypomůže* ... [chybí čárka]  
správně má být: *Tam, kde je to zapotřebí, vypomůže*

s. 40, 2. řádek odshora: *Jendou*, má být jistě *Jednou*

s. 53, 1. řádek shora: *můžu*, má být jistě *mužů*

a to zdaleka není vše...

Závěrem: Práce velmi povrchně a nemoderně pojednává o komunikaci s počítačem v přirozeném jazyce, a to na základě dost elementárních, úvodních a dost starých publikací (starých přes 25 let!). Jako by autorka vůbec nereflektovala, nebo jen marginálně reflektovala např. převratný vývoj v oblasti lingvistické metodologie: *jazykové korpusy*, jejichž využívání dost mění dosavadní pohled na jazyk, a vskutku nepřehledné množství relevantní lingvistické literatury nahromaděné za posledních cca dvacet let. Patrně nezná nějaké novější jazykovědné publikace. Domnívám se, že na práci měla studentka málo času, proto je také práce napsána obsahově i jazykově dost nedbale. Kdybych měl navrhnout známku, tak je to na hraně přijatelnosti: **sotva trojku**. Vzhledem ke stanovení výsledné známky považuji rozpravu při obhajobě za velmi významnou.



V Praze 23. 1. 2010

doc. RNDr. Vladimír Petkevič, CSc.