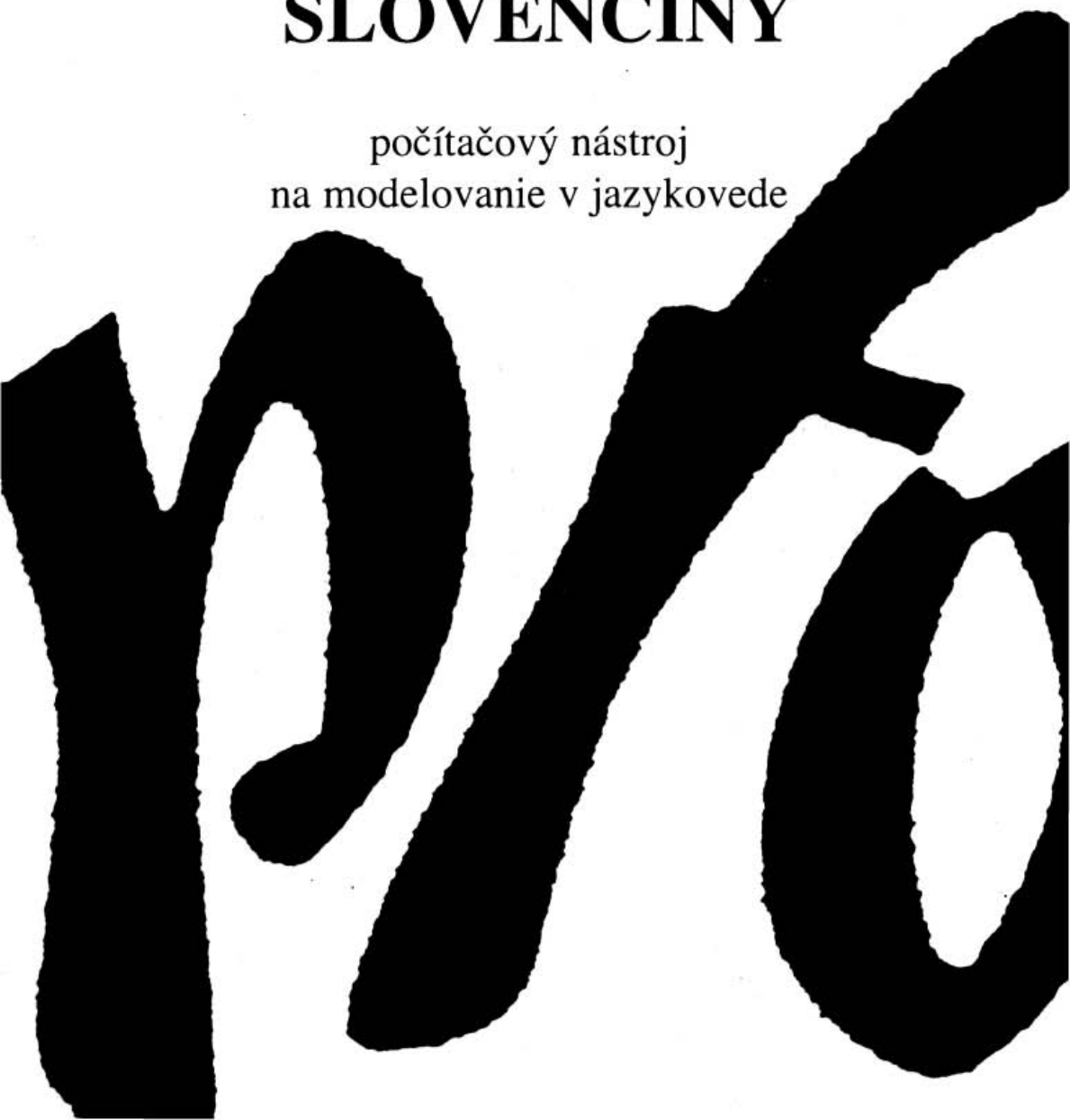


Emil Páleš

*Sapfo*

**PARAFRÁZOVAČ  
SLOVENČINY**

počítačový nástroj  
na modelovanie v jazykovede



**SLOVENSKÁ AKADÉMIA VIED**  
**JAZYKOVEDNÝ ÚSTAV ĽUDOVÍTA ŠTÚRA**

RECENZENTI

Doc. RNDr. Jozef Kelemen, CSc.

Doc. PhDr. Eva Ružičková, CSc.

Doc. PhDr. Juraj Dolník, DrSc.

**Emil Páleš**

*Sapfo*

**PARAFRÁZOVAČ  
SLOVENČINY**

počítačový nástroj  
na modelovanie v jazykovede



VEDA  
VYDAVATEĽSTVO  
SLOVENSKEJ  
AKADÉMIE  
VIED  
BRATISLAVA  
1994

© E. PÁLEŠ 1994

ISBN 80-224-0109-9

## **Venovanie**

*Mojim rodičom,  
a všetkým tým, čo pracovali rukami,  
aby som ja mohol pracovať hlavou;  
mojim univerzitným profesorom,  
ktorí ma naučili myslieť  
a mojim duchovným učiteľom,  
ktorí ma naučili to,  
čo je viac ako myslenie.*

## **Podakovanie**

*PhDr. Kláre Buzássyovej, CSc.*

*Prof. PhDr. Jánovi Horeckému, DrSc.*

*Doc. RNDr. Jozefovi Kelemenovi, CSc.*

*RNDr. Petrovi Mikuleckému, CSc.*

*PhDr. Jarmile Panevovej, DrSc.*

*PhDr. Jane Sejákovej*

*za množstvo rád a pripomienok*

*a anjelom,*

*ktorí mi našepkali to,  
čo je v tejto práci najcennejšie.*

## OBSAH

0.1	O ciele a význame výpočtovej lingvistiky .....	11
0.2	Turingov test.....	14
0.3	Eliza.....	15
0.4	Návrh papagájovacieho stroja .....	16
0.5	Parafráza ako kritérium porozumenia .....	18
0.6	Skutočný kognitívny model .....	18
0.7	Stratifikačná sekvenčná analýza jazyka .....	19
0.8	Rozlišovacia hierarchia .....	22
0.9	Nejednoznačnosť .....	23
0.10	Sekvenčný a paralelný model.....	26
0.11	Synergia. Súčinnosť jazykových podsystémov .....	28
0.12	Konekcionizmus a subsymbolové výpočty .....	29
0.13	Všeobecnosť versus hĺbka .....	30
0.14	Experiment Par Ex (Paraphrasing of Expressions) .....	32

### PRINCÍPY

0.15	Princíp jedno-jednoznačnosti reprezentácie .....	33
0.16	Princíp kontrastívnej diferenciácie.....	34
0.17	Synergický princíp .....	35
0.18	Princíp 51 % .....	36

### SLOVNÍKY

1.1	Údajová a procedurálna časť .....	37
1.2	Morfologický slovník .....	37
1.3	Lexikálno-sémantický slovník .....	47
1.4	Slovník syntaktických väzieb.....	53
1.5	Slovník slovotvorných hniezd .....	54
1.6	Synonymický a lexikálno-paradigmatický slovník .....	61
1.7	Frazeologický slovník .....	64
1.8	Štylistický slovník .....	65
1.9	Slovník asociatívnych väzieb a konotácií .....	66

### FONOLÓGIA

2.1	Klasifikácia foném .....	67
-----	--------------------------	----

### MORFOLÓGIA

3.1	Potreba morfolologickej analýzy.....	69
3.2	Definícia morfolologickej analýzy a syntézy .....	69

3.3	Procedúra analýzy .....	70
3.4	Izolácia koreňa .....	71
3.5	Hypotézy o rozdelení na koreň a afixy .....	71
3.6	Príklad.....	78
3.7	Hypotézy o alternáciách .....	79
3.8	Minimálna nevyhnutná nedeterministickosť .....	80
3.9	Vysvetlivky.....	80

## SYNTAKTICKÁ ANALÝZA

4.1	Definícia syntaktickej analýzy .....	81
4.2	Formálna gramatika.....	82
4.3	Chomského hierarchia .....	83
4.4	Bezkontextové gramatiky (CFG) .....	84
4.5	Syntaktická analýza zhora-nadol.....	84
4.6	Syntaktická analýza zdola-nahor.....	86
4.7	Výhody a nevýhody analýzy zhora-nadol a zdola-nahor.....	88
4.8	Zmiešaná stratégia. Analyzátor zhora-nadol s listinou .....	89
4.9	Gramatika Hornových klauzúl (DCG) .....	89
4.10	Čo robí dobrú gramatiku dobrou gramatikou .....	91
4.11	Reverzibilné gramatiky .....	91
4.12	Prechodová sieť (TN) .....	91
4.13	Rozšírená prechodová sieť (ATN).....	92
4.14	Porovnanie generatívnej kapacity sietí a gramatík.....	96
4.15	Deterministickosť.....	96
4.16	Rozšírené prechodové siete pre slovenský jazyk .....	96
4.17	Porovnanie slovenských a anglických ATN .....	97
4.18	Oslabenie prediktívnej sily .....	103
4.19	(Polo)voľný slovosled slovenčiny .....	104
4.20	Polopredikácie vnútri nominálnych fráz .....	104
4.21	Koordinácie a elipsy .....	105
4.22	Taxonomická vzdialenosť pri riešení výpusťiek.....	107

## SÉMANTICKÁ ANALÝZA

5.1	Sémantická štruktúra vety .....	109
5.2	Rozdiel medzi syntaktickými a sémantickými rolami .....	109
5.3	Sémantické pády.....	110
5.4	K pôvodu sémantických pádov .....	112
5.5	Problém voľby sémantických pádov .....	113
5.6	Sekundárna špecifikácia sémantických pádov .....	113
5.7	Príklady.....	114
5.8	Sémantika slovesných fráz .....	115
5.9	Valencia.....	116
5.10	Intenčné rámce .....	116
5.11	Triedy ekvivalencie slovies .....	117
5.12	Niektoré slovesné intencné rámce.....	119
5.13	Pádové preferencie .....	123



5.14	Sémantika druhotných predložiek. Nešpecifické doplnenia .....	124
5.15	Počítadlo povinnej intencie .....	124
5.16	Sémantika prísloviek .....	125
5.17	Sémantika vedľajších viet.....	125
5.18	Kontrastívny príklad.....	126
5.19	Kontrastívny príklad.....	127
5.20	Protipríklad.....	127
5.21	Sémantika nominálnych fráz.....	128
5.22	Nominalizácia.....	128
5.23	Funkcie LVAL, RVAL .....	129
5.24	Prívlastkové vedľajšie vety.....	131
5.25	Kontrastívny príklad.....	132
5.26	Kontrastívny príklad.....	134
5.27	Protipríklady .....	135
5.28	Súvetná sémantika.....	136
5.29	Príklady.....	136
5.30	Protipríklady .....	137
5.31	Nepravá parataxa .....	137
5.32	Nepravá hypotaxa.....	137
5.33	Sémantika infinitívnych fráz .....	138
5.34	Sémantika adjektívnych fráz .....	138
5.35	Translokácia vetných členov .....	139
5.36	Sémantika doplnku.....	140
5.37	Kontrastívny príklad.....	141
5.38	Združené pomenovania .....	142
5.39	Modálna stavba vety.....	143

## INTEGROVANÁ SYNTAKTICKO-SÉMANTICKÁ ANALÝZA

6.1	Nárast nejednoznačností pri sekvenčnej analýze.....	145
6.2	Paralelná syntakticko-sémantická analýza.....	146
6.3	Predvýpočet valenčných polí.....	147
6.4	Špecifikum flektívnych jazykov .....	149
6.5	Interakcia syntaxe, sémantiky a morfológie. Príklad .....	149
6.6	Principiálne ohraničenia sémantických gramatík .....	152

## DERIVATOLOGICKÁ ANALÝZA

7.1	Potreba slovotvornej analýzy .....	157
7.2	Radikálový rozklad.....	158
7.3	Funkcie CVAL, DVAL .....	159
7.4	Translexikálna syntaktická analýza.....	161
7.5	Prelínanie supra- a sublexikálnej úrovne .....	162
7.6	Viacznačnosť a vágnosť slovotvorného významu .....	163
7.7	Petrifikácia, lexikalizácia, idiomatizácia.....	164
7.8	Slovotvorná parafráza na počítači .....	165
7.9	Synergický efekt pri slovotvornom parafrázovaní.....	166
7.10	Kontrastívny príklad.....	168

7.11	Kontrastívny príklad.....	169
7.12	Kontrastívny príklad.....	171
7.13	Kontrastívny príklad.....	171
7.14	Kontrastívny príklad.....	172
7.15	Kontrastívny príklad.....	173
7.16	Protipríklad.....	174
7.17	Štruktúra kolokačných funkcií .....	175
7.18	Analógia s chémiou.....	176

## KONTEXTOVÁ ANALÝZA

8.1	Príklady koreferencie .....	177
8.2	Zoznam konštituentov .....	178
8.3	Syntaktická forma. Určitosť, zvratnosť, propriálnosť .....	178
8.4	Kategoriálne zhody. Osoba, rod, číslo, pád, životnosť.....	179
8.5	Stupeň aktivácie .....	179
8.6	Inferenčná metóda .....	180
8.7	Algoritmus SAPFO .....	181
8.8	Problémy.....	181

## FRAZEOLÓGIA

9.1	Petrifikované a voľné frazeologické jednotky.....	184
9.2	Reprezentácia ďalších frazeologických javov .....	185

## ŠTYLISTIKA

10.1	Štylistický modul.....	187
10.2	Štylistické prostriedky lexikálne .....	187
10.3	Štylistické prostriedky hláskoslovné.....	188
10.4	Štylistická hodnota gramatických tvarov .....	189
10.5	Štylistické návestia. Dialekty a archaizmy.....	189
10.6	Štylistická hodnota syntaktických konštrukcií.....	190

## STAVEBNICOVÁ REPREZENTÁCIA VÝZNAMU

11.1	Vlastnosti sémantickej reprezentácie .....	191
11.2	Sentence Representation Language (SRL).....	191
11.3	Komponenty významu .....	192

## SYNTÉZA PARAFRÁZ

12.1	Intonácia .....	195
12.2	Transfokácia .....	195
12.3	Transformácia.....	196

12.4	Príklady transformácií .....	197
12.5	Vznik dvojznačných parafráz .....	199
12.6	Transdominácia .....	199
12.7	Konverzia .....	201
12.8	Perspektivizácia .....	203
12.9	Faktické pády .....	204
12.10	Ambivalencia doplnku .....	205
12.11	Selekcia.....	207
12.12	Komplexné vetné členy .....	207
12.13	Transfigurácia.....	209
12.14	Neutralizácia komponentov významu v kontexte .....	210
12.15	Transmutácia .....	211
12.16	Syntaktické, sémantické a faktické manipulácie s jazykom .....	212
12.17	Nesynonymické parafrázovanie I.....	213
12.18	Definícia parafrázy .....	213
12.19	Nesynonymické parafrázovanie II.....	214
12.20	Príklady nesynonymných parafráz .....	214
12.21	Rozdiel medzi trans- formáciou, figuráciou a mutáciou .....	216
12.22	Kovanie čistého významu do foriem.....	216
12.23	Analógia s metalurgiou .....	219
12.24	Veľký príklad.....	220
12.25	Strojová poézia .....	226

## **PREHLAD SYNTAKTICKÝCH, SÉMANTICKÝCH A FAKTICKÝCH MANIPULÁCIÍ**

13.1	Transmutácie gramatické .....	227
13.2	Transmutácie lexikálno-sémantické .....	229
13.3	Transfigurácie .....	232
13.4	Transdominácia .....	234
13.5	Transformácie syntaktické .....	235
13.6	Transformácie kondenzačné.....	236
13.7	Transformácie eliptické .....	240
13.8	Transfokácia .....	241
13.9	Štylizácia .....	241
13.10	Básnické figúry .....	243

## **INTELIGENTNÝ NÁVRAT INFORMÁCIE Z DATABANKY**

14.1	Problém sémantickej relevancie .....	247
14.2	Sapfo ako informačný systém .....	247
14.3	Výpočet sémantickej relevancie. Príklad .....	249

## **FREKVENČNÁ ANALÝZA. PRAVDEPODOBNOSTNÁ SELEKCIA PARAFRÁZ**

15.1	Kombinatorická explózia počtu parafráz .....	251
------	--	-----

15.2	Problém volby .....	251
15.3	Fuzzy logika. Spracovanie neistoty .....	252
15.4	Pravdepodobnostné váhové koeficienty a ich štatistické určovanie .....	252
15.5	Usporiadanie parafráz podľa pravdepodobnosti .....	255

## **VÝPOČTOVÁ ZLOŽITOSŤ ALGORITMOV**

16.1	Zložitosť triedy .....	259
16.2	Zložitosť vyhľadávania v slovníku .....	260
16.3	Zložitosť morfologickej analýzy .....	260
16.4	Zložitosť syntaktickej analýzy .....	261
16.5	Ako súvisia zložitosť syntaxe, morfológie a slovtvorby .....	262
16.6	Suboptimálne riešenia .....	263
16.7	Reálny čas výpočtu .....	263

## **ETICKÉ IMPLIKÁCIE VÝSKUMU V UMELEJ INTELIGENCI**

17.1	Cieľavedomé rozhodovanie v problémovom priestore .....	267
17.2	Poznávanie s ohraničenou časovou a energetickou rezervou ...	268
17.3	Hodnoty štruktúrujú nediferencovanosť .....	269
17.4	Model umelej inteligencie .....	271
17.5	Vznik automatizovaných vzorcov správania .....	273
17.6	Vrodené spúšťacie schémy u zvierat .....	274
17.7	Naučené spúšťacie schémy u človeka .....	275
17.8	Kreativita nad racionalitu alebo srdce nad rozum .....	276
17.9	Zúženie a rozšírenie vedomia .....	276
17.10	Globálna kríza ľudstva .....	278

<b>PRAMENE</b> .....	279
----------------------	-----

<b>SUMMARY</b> .....	287
----------------------	-----

<b>ZUSAMMENFASSUNG</b> .....	289
------------------------------	-----

<b>SAPFO. A HIERARCHIC PARAPHRASING ENGINE</b> .....	291
--	-----

## 0.1 O cielech a význame výpočtovej lingvistiky

Výpočtová lingvistika je interdisciplinárna veda na rozhraní kybernetiky, jazykovedy a kognitívnej psychológie. Jej cieľom je vytvoriť formálny model jazykového vedomia. Motivácia takéhoto úsilia je trojaká:

1. Výpočtová lingvistika je súčasťou umelej inteligencie. Umelá inteligencia (intelektika) je veda, ktorá študuje prvky inteligentného správania živých organizmov ako aj strojov. Zaoberá sa hľadaním hraníc a možností symbolickej, znakovkej reprezentácie poznatkov a procesov ich nadobúdania, udržiavania a využívania (Kelemen, 1992). Oblasť porozumenia prirodzeného jazyka patrí medzi centrálnu odvetvu intelektiky spolu s modelovaním zrakového vnímania (počítačové videnie, pozri Marr, 1982), cieľovo orientovaného spájania vnemov s akciami (robotika, Havel, 1980), myslenia (expertné systémy, Popper - Kelemen, 1989) a i.

*Hypotéza o symbolových systémoch* (Newell - Simon, 1976) hovorí, že systémy spracovávania symbolov majú všetky nevyhnutné a postačujúce predpoklady na realizáciu inteligencie. Úlohou umelej inteligencie je túto hypotézu potvrdiť alebo vyvrátiť, prípadne nájsť hranice jej opodstatnenosti (Kelemen, 1992).

Avšak bez ohľadu na to, či sa Newellova a Simonova hypotéza potvrdí alebo vyvráti, metódy a nástroje, ktoré vznikli na poli intelektiky, sú pre človeka užitočné pri zvládnutí celého radu polomechanických praktických úloh. Ak chceme zvládnuť nastávajúcu informačnú explóziu, účinné metódy inteligentného uchovávaní, vyhľadávania a prenosu informácie budú čoraz žiadanejšie. K tomu si treba uvedomiť, že *prirodzený jazyk* je doteraz *jediným* symbolickým nástrojom umožňujúcim pracovať s informáciou *ľubovoľného* obsahu! Uvedené dva dôvody sú už viac ako dostatočnou motiváciou pre existenciu výpočtovej lingvistiky.

Okrem komerčne aplikovaných výsledkov (pozri ods. 0.13) má takýto medzi-odvetvový výskum inšpiratívny účinok aj na ďalšie odvetvia vedy a oplodňuje ich novými myšlienkami. Spomeňme napríklad, že rané syntaktické výskumy N. Chomského, hoci z istého hľadiska označované za neúspešné, boli inšpirátorom vzniku celého nového a progresívneho odvetvia v matematike - teórie formálnych jazykov.

2. Vstup kybernetiky do jazykovedy a vzájomná výmena poznatkov medzi oboma vednými disciplínami znamenala veľkú inšpiráciu pre obe strany.

Spomeňme niekoľko smerov, v ktorých môže výpočtová technika prispieť k renesancii jazykovedy:

a) Jazykoveda je v osobitnej pozícii tým, že skúma prirodzený jazyk pomocou prirodzeného jazyka. Teda predmet a nástroj jej výskumu splyvajú. To je pozícia metodologicky nebezpečná. Navyše, prirodzený jazyk je iba *náznakový*, nie explicitný vyjadrovací systém, čo spôsobuje nedorozumenia. V matematike sa obdobná situácia vyriešila tým, že sa zaviedlo dôsledné rozlišovanie medzi jazykmi a *meta-jazykom*, pomocou ktorého sa tieto jazyky potom skúmajú. Ako metajazyk lingvistiky sa ponúka formálny jazyk matematickej logiky alebo počítačové jazyky.

b) Ďalšou využiteľnou prednosťou počítača je jeho schopnosť spracúvať veľké kvantá informácie, aké by jazykovedec-človek nikdy nemohol obsiahnuť. Jazykový prejav je fenomén taký zložitý, živý a rozmanitý, že žiaden ľudský subjekt nemôže formulovať svoje hypotézy o jazyku pri plnom vedomí všetkých vzťahov, ktoré môžu s jeho objektom záujmu súvisieť a ovplyvniť ho. U človeka vždy musí nastať *zúženie vedomia koncentráciou*. To znamená, že ak sa výskumník sústreďí na niektorú podoblasť, ostatné oblasti musí čiastočne pustiť z hlavy. Tieto upadajú do mierneho polotieňa a vzťahy v nich sa musia v tej chvíli chápať zjednodušene, schematicky.

Jedine stroj si dokáže naraz pamätať a brať do úvahy všetky zákonitosti všetkých podsystémov jazyka a testovať ich správanie v plnej vzájomnej interakcii. Len počítač dokáže overiť každú hypotézu na mnohomiliónoch korpusoch a oslobodí jazykovedca od sugescie špeciálneho príkladu.

c) Tretím pádnym dôvodom na aplikáciu výpočtovej techniky v jazykovede je, že počítačové modelovanie nadobúda funkciu *exaktnej aplikačnej oblasti*, ktorá lingvistiku stimuluje a koriguje *neúprosňm* spôsobom. Lingvistika doteraz takúto oblasť nemala, aspoň nie v takom zmysle, ako napríklad fyzika. Definitívnym potvrdením či vyvrátením nejakej teoretickej hypotézy v exaktných vedách bol *praktický experiment konštruktívneho charakteru*. Mohlo to byť napríklad skonštruovanie nejakého užitočného zariadenia, ktorého funkčnosť, prípadne zlyhanie, sú potom očividné.

Takýmto testom je pre jazykovedu konštrukcia *parafrázovača* alebo *dialógového systému*.

3. Každý model rečovej komunikácie je zároveň modelom ľudskej mysle. Pri tvorbe kognitívno-lingvistického modelu sa nevyhneme otázke, ako súvisí jazyk a reč s myslením, cítením, vnímaním a konaním človeka. Počítačová lingvistika nám teda v neposlednom rade pomáha - *pochopiť nás samých!*

Každý, aj neúspešný experiment v tomto smere prispieva k prehodnoteniu našich postojov a preformulovaniu otázok zmyslupnejším spôsobom. Keby jediným výsledkom umelej inteligencie malo byť poznanie, že hypotéza o symbolových systémoch *nie je* pravdivá, ale aj poznanie, *prečo* nie je pravdivá, znamenalo by to veľký pokrok ľudského ducha, ktorý by takto pochopil, *v čom spočíva rozdiel medzi strojom a človekom, medzi racionalitou a kreativitou, medzi púhou repetitívnosťou a živým duchom*.

## 0.2 Turingov test

Čo je inteligencia a čo je kritériom jej prítomnosti?

Na rozhraní štyridsiatych a päťdesiatych rokov si túto otázku položil jeden z priekopníkov matematického štúdia počítačov a výpočtov britský matematik Alan M. Turing. Výsledkom bol návrh tzv. *imitačnej hry* (Turing, 1950, skrátene pozri aj v Gál - Kelemen, 1992).

„Imitačnú hru hrajú traja hráči. Hráč A má informácie o hráčoch B a C iba v podobe textov, ktoré vysiela a dostáva napr. cez alfanumerický displej počítačovej siete alebo tak, ako to pôvodne navrhoval Turing, cez ďalekopis. Jeden z hráčov B a C je muž, druhý žena. Hráč A má odhaliť, kto je žena a kto muž. Muž má prekažiť výhru hráča A. Cieľom hry ženy je, aby hráč A vyhral. Podľa Turinga by bol počítač inteligentný, keby v takejto imitačnej hre dokázal úspešne hrať podľa stratégie muža. Takto získala imitačná hra štatút akéhosi testu na zisťovanie prítomnosti inteligencie v počítačových programoch. Do literatúry o umelej inteligencii vošla pod názvom *Turingov test*.“ Podľa Turingovho testu by teda počítač prejavil inteligenciu, ak by vo svojom jazykovom prejave dokázal byť neodlíšiteľný od človeka.

### 0.3 ELIZA

V Shawovej hre *My Fair Lady* urobil profesor fonetiky Higgins z pouličnej kvetinárky Elizu, inak analfabetku, duchaplnú dámu. Urobil to tak, že s ňou namemoval niekoľko anglických fráz (napríklad o počasí) vrátane dôkladnej výslovnosti, ktoré potom použila v určitých typických situáciách. S takouto výzbrojou Eliza obstála v plytkej spoločenskej konverzácii veľmi dobre!

V polovici šesťdesiatych rokov vznikol na Masachussettskom technologickom inštitúte program ELIZA (Weizenbaum, 1966), ktorý sa vzápätí stal jedným z najpopulárnejších programov v umelej inteligencii. Dialógy s Elizou pôsobili tak prirodzene, že viacerí odborníci sa sprvu domnievali, že ELIZA obstála v Turingovom teste:

Človek:	ELIZA:
Ludia sú všetci rovnakí.	V akom smere?
Vždy od teba niečo chcú.	Môžeš mi povedať nejaký príklad?
Tak dobre. Môj chlapec chcel,	Teda tvoj chlapec chcel, aby si
aby som sem prišla.	sem prišla.
Hovorí, že mám depresie.	Mrzí ma, že máš depresie.
Naozaj, som nešťastná.	Myslíš, že príchod sem ti pomôže
	nebyť nešťastnou?
Potrebujem pomoc; to je isté.	Čo by to pre teba znamenalo, keby
	si dostala pomoc?
Možno by som sa naučila lepšie	Povedz mi niečo viac o tvojej ro-
vychádzať s matkou.	dine.
Matka sa o mňa stará.	Kto z rodiny sa o teba ešte sta-
	rá?
Môj otec.	Hmmm, otec.
V niečom sa podobáš na môjho	Akú podobnosť vidíš?
otca.	



Paradoxné bolo, že tento jednoduchý program nemal skoro žiadne vedomosti o jazyku! ELIZA pracovala na základe *klúčových slov a spúšťacích mechanizmov* (trigger configurations). Program pozostával z tabuľky kľúčových slov, preferencií a niekoľkých šikovne volených výstupných frázových vzorcov. Kľúčové slová fungovali ako spúšťače bez ohľadu na kontext. Ku kľúčovému slovu potom ELIZA vybrala niektorú zo štandardných odpovedí s tým, že určité sekvencie prepísala frázami, ktoré počula u svojho partnera. Vzorce striedala, aby sa neopakovali.

<i>klúčové slovo</i>	<i>preferencia</i>	<i>vzorec</i>	<i>výstup</i>
podobný, rovnaký	10	?x	V akom smere? Akú podobnosť vidíš?
je	3	?x je ?y	Bol by si radšej, keby x nebolo y?
	3	?x ?y-uje	Čo keby x ne-y-oval?
vždy	5	?x	Môžeš mi povedať nejaký príklad? Kedy?
čo	2	?x	Prečo sa pýtaš? Zaujíma ťa to?
otec, matka	7	?x	Povedz mi niečo viac o tvojej rodine.

Weizenbaumov program je teda skôr kolekciou vtipných trikov, než kognitívny modelom. Obstál naozaj v Turingovom teste?

Nie. Pri bližšom zoznámení s Elizou by rýchlo vyšlo najavo, aká je plytká. Počas dlhšieho dialógu by sa ukázalo, že rozhovor s Elizou nemá pointu, že je „o ničom“. Program nerozpoznával ani chybné formulované a zmysluprázdne frázy. Takže keby ste povedali nezmysel, Eliza by si to nielen že nevšimla, ale ešte by to po vás aj zopakovala. Navyše, Eliza vystupovala vždy v roli psychoterapeuta a človek v role pacienta. To jej umožňovalo väčšinu času sa pýtať a viesť rozhovor zdanlivo disparátnymi smermi, čo na prvý pohľad pripomína postup psychológa. Každý priamej otázke sa vyhla protiotázkou. Otázky ako „*Prečo sa pýtaš?*“, „*Ako odpovieš na túto otázku?*“ by ju odhalili.

## 0.4 Návrh papagájovacieho stroja

Pozoruhodné a vtipné sú pokusy s tzv. *papagájovacím strojom* (Páleš, 1987). Papagájovací stroj je vlastne *asociatívny generalizátor*. Pracuje takto: pozdrav zopakuje a vyzve človeka na rozhovor. Potom človek a stroj striedavo preberajú iniciatívu. Človek niečo povie, stroj reaguje a naopak.

Papagájovač pritom zovšeobecňuje typy otázok a odpovedí a spája ich asociatívnymi väzbami. Snaží sa zistiť, akým spôsobom sa reaguje na rozličné druhy viet.



Ak narazí na nový typ otázky, odpovedá „neviem“ a otázku si zapamätá. Po nejakom čase sa opýta podobným spôsobom a zapamätá si, akým spôsobom na ňu človek odpovedal. Na oznamovacie vety najprv odpovedá „to je zaujímavé“ a neskôr pri nejakej príležitosti iniciuje rozhovor analogickým spôsobom. Opäť si zapamätá, ako sa na takú vetu reaguje.

Papagájovač teda vytvára dvojice asociovaných vetných vzorcov - akcií a reakcií. Vo vetnom vzorci je zapamätaná gramatická štruktúra, sémantické triedy slov, ako aj prítomnosť slovotvorných afixov a vzťahy zretazovania medzi slovami. Relaxáciou (upustením od) jednotlivých komponentov vzorca potom vznikajú nadvzorce rôzneho stupňa zovšeobecnenia. Keď človek iniciuje rozhovor nejakou akciou, stroj sa snaží unifikovať jej typ s nejakým zovšeobecneným typom akcie v pamäti a pozrie sa, aké typy reakcií k nej patria. Nakoniec naplní reakčný vzorec náhodnými lexémami, ale podľa možnosti tak, aby zachovávali čo najväčší počet komponentov pôvodného vzorca. Papagájovací stroj je pôsobivý tým, že na začiatku nemá žiadne vedomosti, ale v priebehu rozhovoru sa učí. Neskôr potom na svojho spoločníka v konverzácii urobí nečakaný dojem „inteligenciou“, ktorú mu v priebehu predchádzajúceho rozhovoru nenápadne „ukradol“:

- 01 Človek: Ahoj!
- 02 Papagájovač: **Ahoj!**
  
- 03 Papagájovač: Chceš sa so mnou rozprávať?
- 04 Človek: Chcem.
  
- 05 Človek: Ako sa máš?
- 06 Papagájovač: Nevieam.
  
- 07 Papagájovač: Prečo kvitne pšenica?
- 08 Človek: Pšenica predsa nekvitne! Iba kvety kvitnú.
  
- 09 Človek: Zajtra pôjdem pozbierať snežienky.
- 10 Papagájovač: To je zaujímavé.
- 11 Papagájovač: **Ako sa holíš?**
- 12 Človek: Holiacim strojčekom. Alebo žiletkou.
  
- 13 Človek: Prečo sa pýtaš?
- 14 Papagájovač: Ja sa **predsa nepýtam!** Iba pastieri sa pýtajú.
  
- 15 Papagájovač: **Zajtra pôjdem pokosiť stromčeky.**
- 16 Človek: Ako ich chceš *kosiť*?
- 17 Papagájovač: **Kosiacim mlynčekom. Alebo pieckou.**
- 18 Človek: Chceš si zo mňa strieľať?
- 19 Papagájovač: **Chcem.**

Pozorne si všimnite genézu otázok a odpovedí. V piatom riadku ešte papagájovač nevedel odpovedať na otázku, ktorej typ mu bol neznámy. Ale už v riadku 11 sa opýtal tým istým spôsobom, aby zistil, ako sa na takú otázku odpovedá. Svoj nový poznatok potom úspešne použil v riadku 17. Analogicky k (12) dodržal zhodu slovesa so slovesom v otázke a substantíva vybral z rovnakej lexikálnej skupiny. Podaril sa mu dokonca taký výber lexiky, ktorý aj zvukovo kopíruje pôvodnú tvarovú a slovotvornú štruktúru.

Iniciáciu rozhovoru (15) prevzal z riadku 9, kde ju počul po prvý raz. Dvojica 13, 14 je zovšeobecnením 7, 8 a 3, 4 sa opakuje v 18, 19.

## 0.5 Parafráza ako kritérium porozumenia

Ak učiteľ chce zistiť, ako žiak pochopil látku, požiada ho, aby ju vysvetlil vlastnými slovami. Zo žiakovej odpovede je potom očividné, koľko z obsahu skutočne porozumel a čo iba memoruje.

Práve tak aj počítač svojou (ne)schopnosťou parafrázovať vety dáva zreteľne najavo, do akej miery a v akých smeroch pochopil, či nepochopil ich význam.

Úlohou parafrázovača je generovať práve množinu všetkých parafráz danej frázy; nič viac ani nič menej (porov. Chomsky, 1956). Inými slovami: maximalizovať množstvo správne vygenerovaných parafráz a súčasne minimalizovať množstvo parafráz nesprávnych. Za parafrázu pritom pokladáme každú frázu, ktorá je s pôvodnou frázou významovo usúvzťažnená.

Úlohu parafrázovača pokladáme za ekvivalentnú s Turingovým testom. Ak by totiž existoval počítač, ktorý by správne parafrázoval všetky frázy vo všetkých kontextoch, nebol by vo svojom jazykovom prejave odlíšiteľný od človeka.

Ako alternatívne a ekvivalentné kritérium môže slúžiť aj *schopnosť* počítača viesť dialóg.

## 0.6 Skutočný kognitívny model

Žiaden trikový systém zrejme neobstojí a nemôže obstáť ani v Turingovom teste, ani ako parafrázovač či dialógový systém. Jediný nádejný smer je pokúsiť sa vybudovať model, ktorý by bol skutočným odrazom kognitívnych štruktúr jazykového vedomia človeka. Tu stoja pred nami dva centrálné problémy (Allen, 1987):

1. *Problém reprezentácie*: Ako majú tieto štruktúry, ktoré reprezentujú obsah jazykového vedomia, vyzeráť?

2. *Problém interpretácie*: Ako túto reprezentáciu algoritmicke vyrobiť? T. j. ako vyzerá konštruktívny návod, pomocou ktorého môžeme transformovať povrchové vetné štruktúry na hĺbkovú reprezentáciu a naopak?

Najzaužívanejším formalizmom pre reprezentáciu významu sú rámce (Minsky, 1975) alebo sémantické siete (napr. Winston, 1977).

Klasickým prístupom k riešeniu problému interpretácie je stratifikačná sekvenčná analýza jazyka. Vstupný reťazec (veta) pri nej postupne prechádza fonó-

logickou, morfológickou, syntaktickou, sémantickou a pragmatickou *analýzou*, ktorých výsledkom je hĺbková reprezentácia. *Syntéza*, t. j. premena hĺbkových štruktúr späť na lineárne reťazce jazyka, prebieha tými istými úrovňami v opačnom poradí (pozri obrázok 1).

## 0.7 Stratifikačná sekvenčná analýza jazyka

### Fonologická analýza

V systémoch s rečovým vstupom musí hláskoslovná analýza najprv identifikovať vo zvukovom zázname jednotlivé hlásky a slová.

Výsledkom je veta reprezentovaná ako lineárny reťazec znakov abecedy:

*Turecký paša chce napadnúť Starhrad.*

### Morfologická analýza

Tvaroslovná analýza identifikuje gramatické kategórie jednotlivých slov vo vete.

Výsledkom je zoznam gramatických kategórií ku každému slovu:

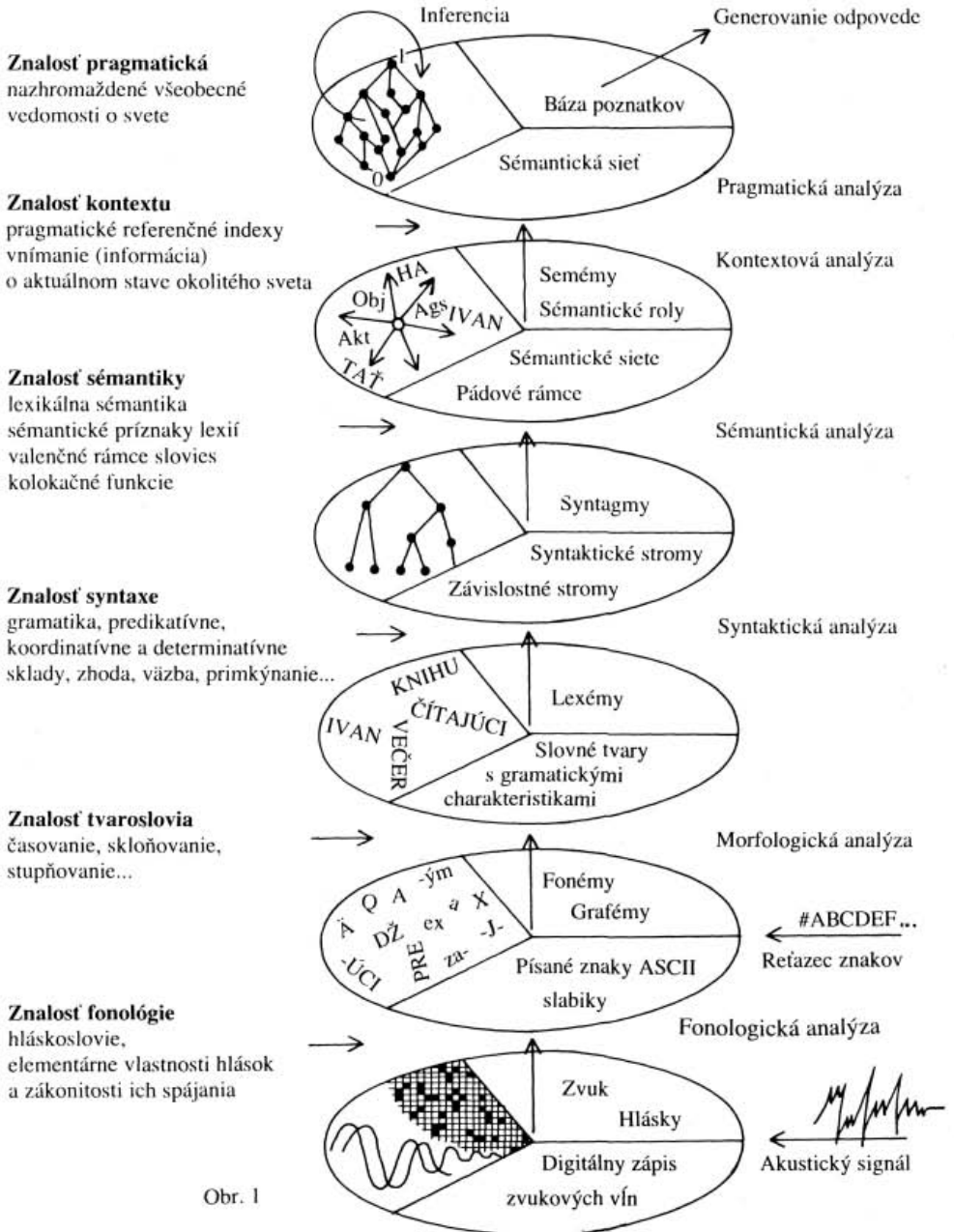
<i>turecký</i>	adj.	muž.	živ.	nom.	sg.	poz.	
	»	»	než.	nom.	«	«	
	»	»	než.	ak.	«	«	
<i>paša<sub>1</sub></i>	sub.	muž.	živ.	nom.	sg.		(vládár)
<i>paša<sub>2</sub></i>	sub.	žen.	než.	nom.	sg.		(pastva)
<i>chce</i>	pr.	indik.	3.os.	sg.	prít.čas.	poz.	
<i>napadnúť<sub>1</sub></i>	infin.						(zaútočiť)
<i>napadnúť<sub>2</sub></i>	infin.						(sneh)
<i>napadnúť<sub>3</sub></i>	infin.						(prísť na um)
<i>Starhrad</i>	sub.	muž.	než.	nom.	sg.		
	»	»	»	ak.	sg.		

### Syntaktická analýza

Syntaktická analýza identifikuje vetné členy, ich skladbu do fráz (syntagiem) a popri ňom im tzv. syntaktické roly (podmet, prísudok, predmet, nepriamy predmet, prívlastok, prístavok, doplnok...).

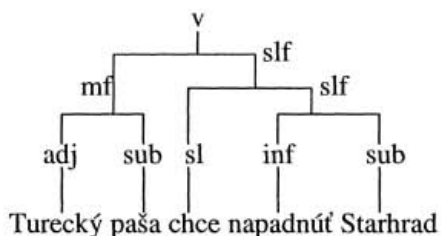
Výsledkom je syntaktická štruktúra vyjadrujúca vzťahy formálnej závislosti vo vete. Reprezentuje sa uzátvorkovaným zoznamom alebo derivačným stromom:

## Sekvenčný stratifikačný model jazykovej analýzy



Obr. 1

chcieť(paša(turecký),napadnúť(Starhrad))

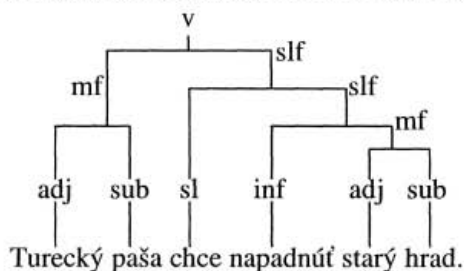


### Morfematická analýza

Morfematická analýza identifikuje *slovotvornú* štruktúru, t. j. vzťahy medzi viacnásobnými koreňmi a afixami vnútri zložených a odvodených slov.

Reprezentácia je analogická syntaktickej a je jej rozšírením:

chcieť(paša(turecký),napadnúť(hrad(starý)))

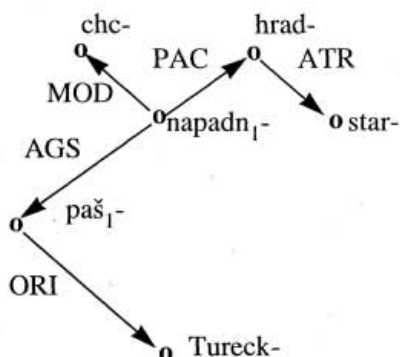


### Sémantická analýza

Vykonať sémantickú analýzu znamená vo vete identifikovať jednotlivé *významové* participanty (aktanty) a priradiť im príslušnú sémantickú rolu (pád).

Výsledkom je sémantická štruktúra vyjadrujúca vzťahy *významovej* závislosti. Reprezentuje sa pádovými rámcami alebo sémantickou sieťou:

napadnúť<sub>1</sub>  
 AGENS: paša<sub>1</sub>  
 ORIGINATÍV: Turecko  
 PACIENS: hrad  
 ATRIBÚT: starý  
 MODÁL: chcieť  
 ČAS: prítomný



## Kontextová analýza

Kontextová analýza uvádza vetu do súvislosti s kontextom. Identifikuje súvislosti medzi členmi, ktoré sa nenachádzajú v tej istej vete, ale sa na seba vzťahujú.

Dajme tomu, že po vete *Turecký paša chce napadnúť Starhrad* nasleduje veta:

*Knieža Ctibor sa pripravuje na obranu proti Sulejmanovi III.*

Potom úlohou kontextovej analýzy je identifikovať *Tureckého pašu* ako Sulejmana III.; *hrad* z prvej vety ako Ctiborov hrad, Ctiborovu *obranu* ako konverzný dejový protipól k vyššie spomínanému Sulejmanovmu *útoku*, prípadne aktualizovať podľa kontextu čas udalosti, hore špecifikovaný iba ako *prítomný*, na konkrétne dátum. Výsledkom sú takéto kontextové rovnice:

paša = Sulejman III.	(odkazovanie)
hrad = Ctiborov hrad	(zrežazenie)
obrana ≈ napadnutie	(konverzia)
čas = a. d. 1526	(aktuálny čas)

## Inferencia

Úlohou inferencie je efektívne začleniť novú informáciu do už existujúcej poznatkovej siete. Z novej informácie sa snaží vyvodiť logické závery na pozadí všeobecne platných poznatkov alebo poznatkov skoršie získaných. Takto vytvára nové poznatky, resp. rozhoduje posledné nejednoznačnosti, ktoré vo vete zostali. Príklad:

- |  |                      |
|--|----------------------|
| 1. Silné hrady sa dobývali vyhladovaním.   | (všeobecný poznatok) |
| 2. Ctibor sa dobre opevnil.                | (aktuálny poznatok)  |
| 3. Sulejman sa pokúsi Starhrad vyhladovať. | (logický záver)      |
| 4. Ctibor sa zásobil potravinami.          | (aktuálny poznatok)  |
| 5. Sulejman prehrá.                        | (logický záver)      |

## 0.8 Rozlišovacia hierarchia

(1) Zelené žaby žijú v rybníku.	dobře	
(2) Červené žaby majú dlhé nosy.	zle	# pragmaticky
(3) Červené idey majú dlhé nosy.		# sémanticky
(4) Červenými ideou mať nos dlhý.		# syntakticky
(5) Červenskami ideova matý nosník přídlžie.		# morfematicky
(6) Čřvskmn iděe math noss d'lha.		# fonologicky

*Rozlišovacia hierarchia dobre a zle formulovaných viet na rozličných úrovniach.*

Z viet (1) - (6) by sme len (1) prijali bez výhrad. O pravdivosti (2) sa dá diskutovať, o pravdivosti (3) nie. (3) je zle sformulovaná sémanticky, hoci syntakticky je v poriadku. (4) je zle formulovaná syntakticky, ale obsahuje slovenské slová. (5) zvukovo pripomína slovenčinu, ale (6) nie. (5) obsahuje slovenské korene a afixy skombinované do neslovenských slov. Hlásky a spoluhláskové skupiny v (6) už nepatria ani do inventára slovenčiny.

## 0.9 Nejednoznačnosť

Najväčšou prekážkou pri programovaní systémov analýzy prirodzeného jazyka sú nejednoznačnosti. Výsledok analýzy v jednotlivých vrstvách jazyka býva viacznačný a počítač musí pracovať so všetkými alternatívami, až kým nezistí, ktorá z nich sa ukáže ako tá správna. Nezávislé alternatívne možnosti sa medzi sebou násobia, čím vzniká nebezpečie exponenciálnej explózie počtu nejednoznačností a problém nájdenia efektívnych algoritmov, ktoré by pracovali v reálnom čase.

Ilustrujme jednotlivé druhy nejednoznačností na príkladoch:

### *Fonologická nejednoznačnosť*

Na úrovni hláskoslovía vzniká nejednoznačné priradenie medzi zvukmi a fonémami. Jedna fonéma býva fyzicky realizovaná celou škálou rozličných frekvencií a naopak, podobné zvukové vibrácie môžu predstavovať rôzne hlásky. Napríklad v anglických slovách *love*, *book*, *oregon* sa graféma *o* vyslovuje ako naše *a*, *u*, *o* v závislosti od svojho fonologického okolia.

Realizácia fonémy navyše závisí od takých faktorov, ako sú pohlavie, vek, intenzita hlasu, citové zafarbenie a i.

### *Tvarová homonymia*

Jeden tvar môže patriť zároveň k viacerým slovným druhom alebo mu prislúchajú alternatívne súbory (n-tice) gramatických kategórií:

<i>vila</i> :	substantívum/sloveso
<i>ved'</i> :	časťica/imperatív
<i>kvapkám</i> :	sloveso/datív substantíva
<i>hrad</i> :	nominatív/akuzatív
<i>pekných</i> :	gen/ak/lok, muž/žen/str, živ/než

V analytických jazykoch trpí homonymiou značná časť slov. Napríklad slovo *love* je podstatným menom, prídavným menom aj slovesom; *still* dokonca aj prí-slovkou.

### *Lexikálna homonymia*

Jedno slovo (designátor) označuje viac významov (denotátov):

<i>oko</i> :	ľudské, morské, reťaze
<i>koruna</i> :	stromu, peniaz, kráľovská, krajina
<i>krídlo</i> :	dvier, lietadla, vtáčie, futbalové, klavír, skupina ľudí

Homonymia sa vylúči až počas syntaktickej analýzy v kontexte ostatných slov vo vete: *Anka vila vence*. Alebo selekčným efektom: *Oko a ucho*.

### **Polysémia**

Jeden designátor označuje viacero denotátov (alebo ich odtienkov), medzi ktorými je vnútorný vecný alebo vonkajší formálny súvis: pohľad do slovníka nás presvedčí, že polysémia je systematickým princípom organizácie lexiky:

<i>padnúť</i> :	padnúť <sub>1</sub> do studne (klesať)
	padnúť <sub>2</sub> za vlasť (zomrieť)
	pevnosť padla <sub>3</sub> (bola dobytá)
	padol <sub>4</sub> návrh (odznel)
	lós padol <sub>5</sub> na mňa (vyšiel)
	padnúť <sub>6</sub> do pasce (ocitnúť sa)
	šaty jej padnú <sub>7</sub> (pristanú)
...	padnúť <sub>8</sub> niekomu do oka (zapáčiť sa)
	dvanásť hodina padla <sub>9</sub> (odbila)
	padla <sub>10</sub> kosa na kameň

Polysémiu sčasti rieši sémantická analýza a síce silou selekčného efektu sémantických kolokácií:

<i>zlatý</i> <sub>1</sub>	<i>prsteň</i>
<i>zlaté</i> <sub>2</sub>	<i>dievča</i>

### **Syntaktická viacznačnosť**

Jedna veta či fráza umožňuje viac spôsobov syntaktického členenia. České *písek se solí* má dve čítania: *písek sa solí* alebo *písek so solí*. Dve čítania má aj anglická veta:

<i>He</i>	<i>saw</i>	<i>that</i>	<i>gasolin</i>	<i>can</i>	<i>explode.</i>
<i>On</i>	<i>videl</i>	<i>že</i>	<i>petrolej</i>	<i>môže</i>	<i>vybuchnúť.</i>
<i>»</i>	<i>»</i>	<i>ten</i>	<i>petrolejový</i>	<i>kanister</i>	<i>«</i>

### **Kontextová nejednoznačnosť**

Slová vo vete môžu odkazovať na frázy, ktoré v kontexte predchádzajú (anafora) alebo nasledujú (katafora), alebo ukazovať na objekty z vonkajšieho sveta (exo-



fora). To platí hlavne pre neplnovýznamové (zámená), ale aj pre plnovýznamové slovné druhy.

*Jano zbadal Fera. Pozdravil ho.* (Anaforické odkazovanie. Pozdravil Jano Fera alebo naopak?)

*Hej ty, podaj mi tamto, odtiaľ!* (Exoforické ukazovanie na konsituáciu, za zámená treba doplniť konkrétne entity.)

*Bolo to na príkaz riaditeľa.* (Treba zistiť, kto bol v danom čase na danom mieste riaditeľom.)

### Metaforickosť, metonymickosť

Čítal som (básne od) Hviezdoslava.	metonymia
Dokončil som (písanie) knihu.	asociácia
Na kolokviu bolo veľa múdrych hláv (ľudí).	synekdocha
Hviezdy žmurkajú (trbliecu sa).	metafora

### Intonácia

Odlíšná intonácia a pozmenená kadencia môžu celkom zmeniť význam frázy, alebo zmeniť množinu presupozícií, ktoré z vety nepriamo vyplývajú:

Pekný vtáčik.	Jano prišiel.	(Teda nie Ivan.)
Pekný vtáčik!	Jano prišiel.	(Tak predsa.)

### Pragmatická nejednoznačnosť

Po ukončení rýdzo jazykovej analýzy spravidla zostáva množina nejednoznačností rozhodnuteľných iba aplikovaním znalostí o svete. Podiel týchto nejednoznačností stúpa so stupňom všeobecnosti systému. V úzko špecializovaných systémoch s vyčerpávajúcim popisom domény diskurzu je možné tento druh viacznačností minimalizovať. Ich odstránenie vyžaduje inferenciu s bázou poznatkov:

*fenylhydrazónové, čerpadlo* = 1. čerpadlo z fenylhydrazónu<sub>MAT</sub>  
2. poháňané fenylhydrazónom<sub>INS</sub>  
3. čerpá fenylhydrazón<sub>PAC</sub>

### Komunikatívny zámer

Každá výpoveď je zároveň aj rečovým aktom (Austin, 1962) a implikuje závery o intenciách hovoriacich, ktoré bývajú vlastným zmyslom komunikácie. Pri komunikácii nejde iba o suchopárny prenos informácie, ale účastníci v nej vstupujú do záväzkov, s ktorými potom ostatní počítajú (Winograd - Flores, 1987):