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.

Poďakovanie

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 cieli 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 morfologickej analýzy 69
3.2 Definícia morfologickej 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 výpustiek 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 intečné 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 adjektiválnych 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

PREHĽAD 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 voľby 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 Zložitostné 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 slovotvorby 262
16.6 Suboptimálne riešenia 263
16.7 Reálny čas výpočtu 263

ETICKÉ IMPLIKÁCIE VÝSKUMU V UMELEJ INTELIGENCII

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

PRAMENE 279

SUMMARY 287

ZUSAMMENFASSUNG 289

SAPFO. A HIERARCHIC PARAPHRASING ENGINE 291

0.1 O cieli 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, znakovej reprezentácie poznatkov a procesov ich nadobúdania, udržovania a využívania (Kelemen, 1992). Oblasť porozumenia prirodzeného jazyka patrí medzi centrálne odvetvia 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 uchovávania, 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 medziodvetvový 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 splývajú. 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 *metajazykom*, 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 sústredí 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 mnohomiliónových korpusoch a oslobodiť 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 *neúprosným* 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 zmysluplnejším 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 prekaziť 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 Elizy, inak analfabetky, duchaplnú dámu. Urobil to tak, že s ňou namemoroval 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:*
Ľudia sú všetci rovnakí.	V akom smere?
Vždy od teba niečo chcú.	Môžeš mi povedať nejaký príklad?
Tak dobre. Môj chlapec chcel, aby som sem prišla.	Teda tvoj chlapec chcel, aby si 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 vychádzať s matkou.	Povedz mi niečo viac o tvojej rodine.
Matka sa o mňa stará.	Kto z rodiny sa o teba ešte stará?
Môj otec.	Hmmm, otec.
V niečom sa podobáš na môjho otca.	Akú podobnosť vidíš?

Paradoxné bolo, že tento jednoduchý program nemal skoro žiadne vedomosti o jazyku! ELIZA pracovala na základe *kľúčových 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.

kľúčové slovo	*prefe-rencia*	*vzorec*	*výstup*
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 kognitívnym 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 chybne 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. Každej 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 zreťazenia 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: Neviem.

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 *pýt*aš?
14 Papagájovač: Ja sa **predsa ne***pýt*am! **Iba** pastieri sa *pýt*aj**ú**.

15 Papagájovač: **Zajtra pôjdem po**kosiť stromčeky.
16 Človek: Ako ich chceš *kos*iť?
17 Papagájovač: *Kos***iacim** mlyn**čekom**. **Alebo** piec**kou**.
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 centrálne problémy (Allen, 1987):

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

2. *Problém interpretácie:* Ako túto reprezentáciu algoritmicky 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 fono-

logickou, morfologickou, 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:

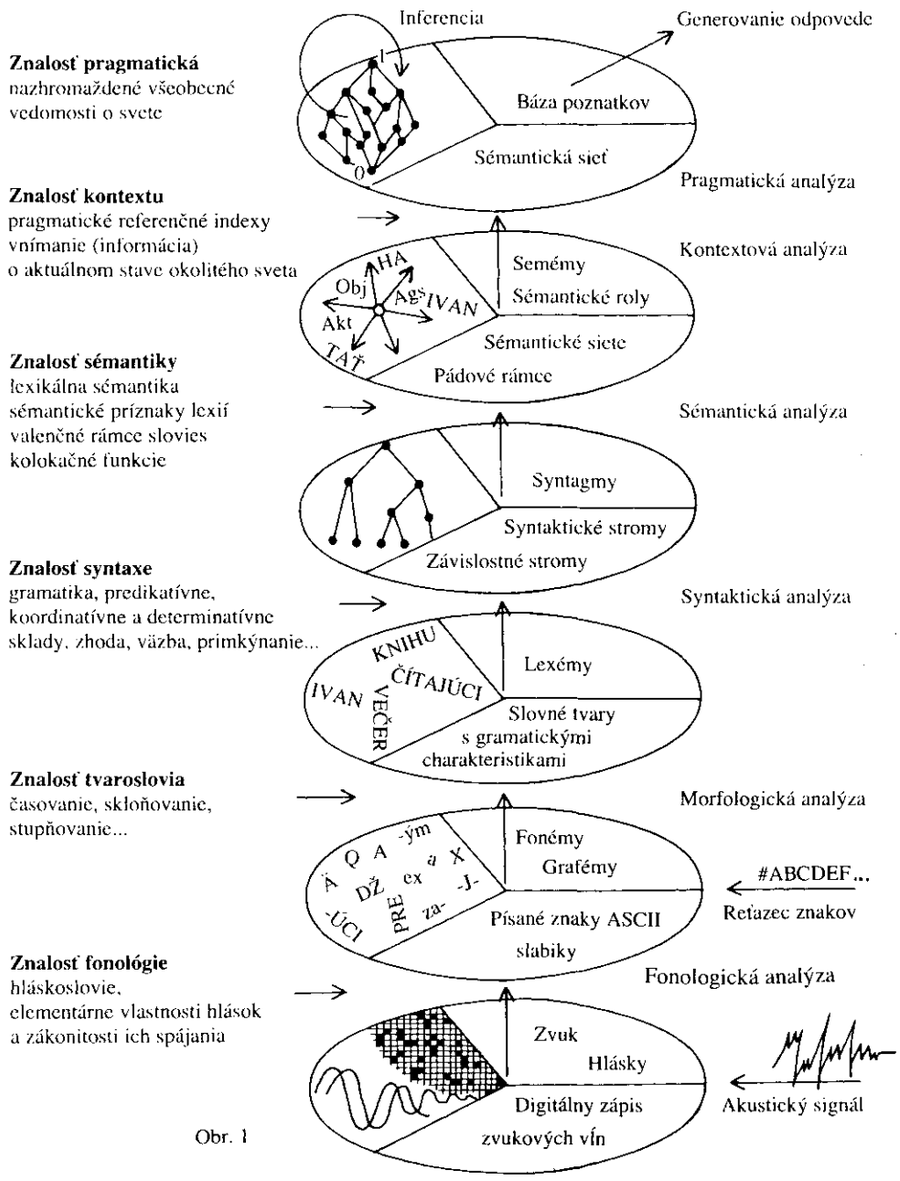
turecký	adj.	muž.	živ.	nom.	sg.	poz.		
	»	»	než.	nom.	«	«		
	»	»	než.	ak.	«	«		
paša$_1$	sub.	muž.	živ.	nom.	sg.			(vladár)
paša$_2$	sub.	žen.	než.	nom.	sg.			(pastva)
chce	pr.	indik.	3.os.	sg.	prít.čas.		poz.	
napadnúť$_1$	infin.							(zaútočiť)
napadnúť$_2$	infin.							(sneh)
napadnúť$_3$	infin.							(prísť na um)
Starhrad	sub.	muž.	než.	nom.	sg.			
	»	»	»	ak.	sg.			

Syntaktická analýza

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

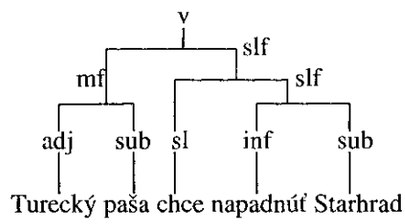
Výsledkom je syntaktická štruktúra vyjadrujúca vzťahy formálnej závislosti vo vete. Reprezentuje sa uzátvorkovaným zoznamom alebo derivačným stromom:

Sekvenčný stratifikačný model jazykovej analýzy



Obr. 1

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

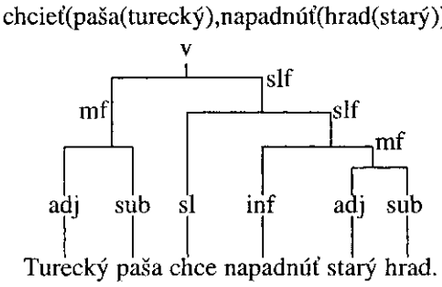


Morfematická analýza

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)))



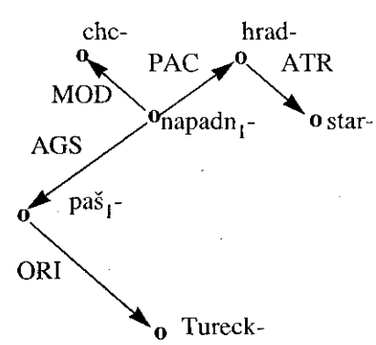
Sémantická analýza

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$_1$
AGENS: paša$_1$
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	(zreťazenie)
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.	dobre	
(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 prídlžie.		# 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 hláskoslovia 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:

vila:	substantívum/sloveso
veď:	častica/imperatív
kvapkám:	sloveso/datív substantíva
hrad:	nominatív/akuzatív
pekných:	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 príslovkou.

Lexikálna homonymia

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

oko:	ľudské, morské, reťaze
koruna:	stromu, peniaz, kráľovská, krajina
krídlo:	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:

padnúť:	padnúť$_1$ do studne (klesať)
	padnúť$_2$ za vlasť (zomrieť)
	pevnosť padla$_3$ (bola dobytá)
	padol$_4$ návrh (odznel)
	lós padol$_5$ na mňa (vyšiel)
	padnúť$_6$ do pasce (ocitnúť sa)
	šaty jej padnú$_7$ (pristanú)
...	padnúť$_8$ niekomu do oka (zapáčiť sa)
	dvanásta hodina padla$_9$ (odbila)
	padla$_{10}$ kosa na kameň

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

zlatý$_1$ *prsteň*
zlaté$_2$ *dievča*

Syntaktická viacznačnosť

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

He	*saw*	*that*	*gasolin*	*can*	*explode*.
On	videl	že	petrolej	môže	vybuchnúť.
»	»	ten	petrolejový	kanister	«

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

Odlišná 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$_{MAT}$
2. poháňané fenylhydrazónom$_{INS}$
3. čerpá fenylhydrazón$_{PAC}$

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):

Na vetu *Prepáčte, stojíte mi na nohe* nemožno odpovedať *Áno pane, je to tak*, pretože z hľadiska komunikatívneho zámeru ide o imperatív, nie o deklaratív. Podobne veta *Stojíš tam ako drevo* je v skutočnosti výzvou k činnosti v danej situácii. Napríklad pri vešaní obrazu môže znamenať: *Podaj mi kladivo!*

Vety *Som lekár* alebo *Nemám rád rasizmus* neobsahujú len púhu informáciu o vzdelaní hovoriaceho, resp. o jeho vkuse či pocitoch. Podstatnou časťou ich *zmyslu* je záväzok, prísľub, že sme kompetentní urobiť zdravotný zásah, resp., že sa zachováme v určitej situácii určitým spôsobom.

Práca s rečovými aktmi vyžaduje modelovanie intencií a presvedčení hovoriacich (napr. Wahlster, 1991).

0.10 Sekvenčný a paralelný model

Videli sme, že sekvenčné moduly si medzi sebou posúvajú a odovzdávajú nejednoznačnosti. Všetko, čo nemožno rozhodnúť na danej úrovni, sa musí zatiaľ odsunúť ako nerozhodnuté a ponechať pre analýzu na niektorej z ďalších úrovní. Dovtedy treba pracovať so všetkými možnosťami ako s plnohodnotnými alternatívami, hoci väčšina z nich je vopred odsúdená na neúspech.

Napríklad vety

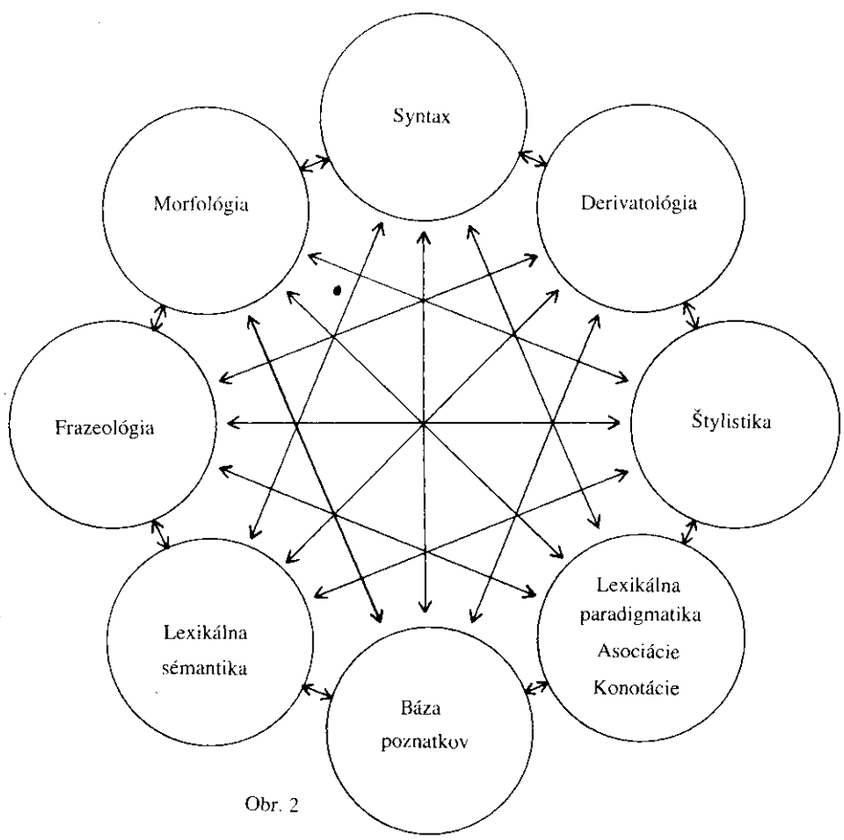
Ivan rozbil vázu zo skla.
Ivan rozbil vázu z hnevu.

sa dajú syntakticky štruktúrovať dvojakým spôsobom. Predložková fráza môže patriť pod predmet alebo pod sloveso. V prvom prípade je podradená predmetu (*váza zo skla*) a v druhom slovesu (*rozbiť z hnevu*). Toto však nie je možné rozhodnúť na úrovni čistej syntaxe, pretože obe vety sú zo syntaktickej stránky rovnorodé. Dištinktívny je sémantický príznak substantíva *sklo* (materiál) a *hnev* (abstraktné). Sémantická analýza sa však v sekvenčnom modeli aktivuje *až po skončení* syntaktickej analýzy, a teda jej nemôže byť nápomocná. Syntaktický modul je odkázaný sám na seba. Musí pracovať s oboma alternatívnymi syntaktickými stromami až do konca, hoci jeden z nich je očividne nesprávny. Počet všetkých nejednoznačností vo vete sa tým zdvojnásobí.

V čom spočíva efektívnosť ľudskej mysle pri spracúvaní jazyka? Predpokladáme, že myseľ spracúva jazyk *celistvo*. Analýza tvarov, skladby a významu prebieha *paralelne* a irelevantné významy sa vylučujú *priebežne*. Naopak. Človek v podstate ani nie je schopný vnímať vetnú skladbu oddelene, bez toho, že by si uvedomoval slovný význam a okamžite inklinoval k preferencii niektorej zo syntaktických štruktúr.

Syntaktický modul by si teda ušetril značné množstvo práce, keby mal v bode nejednoznačnosti *možnosť konzultovať* sémantiku.*

* Komunikácia a dekompozícia, ktorá je jej podmienkou, je užitočná už i v rovine syntaxe, ako to (v kontexte matematickej teórie formálnych jazykov a gramatík) dokazujú E. Csuhaj-Varjú - J. Kelemen



Obr. 2

Holistický (celostný) model analýzy prirodzeného jazyka

Pozostáva z paralelných a voľne komunikujúcich programových modulov. Tokom informácie medzi modulmi vzniká synergický efekt.

Možnosť voľnej komunikácie medzi jednotlivými subsystémami jazyka je základom *holistického* (celostného) modelu jazykovej analýzy (obr. 2). Systém SAPFO pozostáva z ôsmich programových modulov, ktoré stelesňujú jednotlivé jazykové podsystémy a ktoré počas výpočtu medzi sebou komunikujú vždy, keď je to potrebné.

SAPFO teda nie je homogénny konekcionistický model. Pozostáva zo samostatne vyhranených, špecializovaných, ale úzko komunikujúcich modulov. Tým sa darí zachovať žiadúcu modularitu (a s ňou aj prehľadnosť) a zároveň využiť výhodu konekčných modelov.

Existuje *psycholingvistická afázia*, ktorá vyraďuje syntaktickú kompetenciu, ale pritom ponecháva kompetenciu sémantickú. Pacient s takouto afáziou rozumie vety ako *The lion ate up the sheep* ale nerozumie vetu *The lion ate up the tiger*. Prvá veta je totiž zrozumiteľná aj ako neusporiadaný balík lexikálnych jednotiek (*lev,zjesť,ovca)*, ale druhá by mohla znieť aj *The lion was eaten up by the tiger*, čiže syntax v nej *je* relevantná. Z toho usudzujeme na úzku prepojenosť, ale zároveň relatívnu samostatnosť syntaxe a sémantiky u človeka.

0.11 Synergia. Súčinnosť jazykových podsystémov

Synergia (gréc.) znamená spolupôsobenie, *synergický* je spolupôsobiaci, súčinný. *Synergickým efektom* sa rozumie zvýšenie účinku spolupôsobením, alebo dosiahnutie účinku, ktorý prevyšuje sumu účinkov jednotlivých častí samých osebe.

Systém SAPFO je experimentom, či sa v holistickom modeli analýzy a syntézy jazyka uplatní synergický efekt. Analýza algoritmov v nasledujúcich paragrafoch, ako aj merania reálneho výpočtového času (16.7) ukazujú, že synergický efekt naozaj vzniká. Komunikujúce moduly vykonávajú menšie množstvo práce, než aké by vykonávali izolovane. Vzájomná informovanosť im umožňuje včas upustiť od skúmania celého radu bezperspektívnych alternatív.

Napríklad pri analýze vety *Z holí sa ozýva* syntax po prechode predložky *z* už morfológiu nemusí žiadať o analýzu slova *holí* ako slovesa. Tretia osoba od slovesa *holiť sa* by inak bola možná, ale nie po predložke.

Veta *Dub takýchto rozmerov váha nemôže odvážiť* zmätie každý syntaktický analyzátor zľava-doprava a zhora-nadol, pretože jej začiatočná časť je sama vetnou formou: *Dub takýchto rozmerov váha*. Neúspešnosť analýzy vysvitne až pri piatom slove, keď analyzátor zlyhá pri pokuse zaradiť určité sloveso *nemôže* pod určité sloveso *váha*.

Integrovaná syntax-sémantika SAPFO túto zavádzajúcu analýzu ani nezačne. Hneď pri prvom slove sa totiž zistí, že v sémantickom poli intranzitívneho slovesa *váhať*, vyžadujúceho životný subjekt, nemožno neživotný *dub* interpretovať. Začne

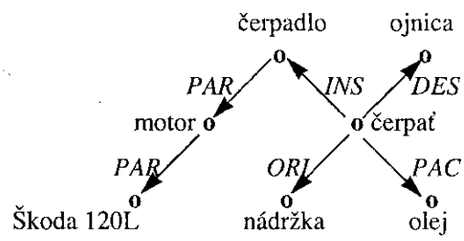
(1991). J. Kelemen (1993) ukazuje niektoré vzťahy takýchto syntaktických výskumov v oblasti formálnych jazykov k súčasnému trendu dekompozície a reaktivity v robotike.

sa teda hneď s analýzou vo valenčnom poli *odvážiť*, ktoré sa tak identifikuje ako centrálne sloveso.

Podrobný príklad vzájomnej podpory a efektívnej spolupráce syntaxe so sémantikou je v odseku 6.5.

V kapitole 7 je rozobraná synergická interakcia slovotvorby s ostatnými podsystémami jazyka. Okrem syntaxe a sémantiky aj lexikálna paradigmatika, konotácie, asociácie, štylistika a frazeológia zasahujú do slovotvornej analýzy rozhodujúcim spôsobom; často by bez nich ani nebola možná.

K synergickému efektu v neposlednom rade prispieva svojou informáciou aj inferenčný modul. Napríklad v poradenskom systéme asistujúcom pri oprave auta môže inak trojznačná fráza *olejové čerpadlo* (výrobný materiál, objekt čerpania, pohonná látka) bez odkladu unifikovať s časťou poznatkovej siete:



Tým fráza postúpi do ďalšej analýzy bez viacznačnosti:

čerpadl-
«INS čerp-
PAC» olej-

Holistický analyzátor sa podobá zhromaždeniu expertov, z ktorých každý je špecializovaný vo svojom odbore. Zišli sa pri riešení spoločného problému, ktorý by nikto z nich nemohol rozriešiť samostatne, pretože sám nemá k dispozícii dostatočnú informáciu.

Sekvenčný analyzátor sa podobá úzko špecializovaným expertom, ktorí riešia interdisciplinárny problém, ale komunikovať medzi sebou môžu až keď sú s prácou hotoví. Posledný expert potom musí preštudovať a vytriediť nepreberné množstvo planých dohadov, ktoré mu postúpili jeho kolegovia.

Robiť syntaktickú, sémantickú a slovotvornú analýzu izolovane znamená vyslať do divočiny troch bádateľov, z ktorých jeden bude mať vodu, druhý nôž a tretí zápalky - ale cestovať musia každý samostatne.

0.12 Konekcionizmus a subsymbolové výpočty

Myšlienky paralelizmu a konekcionizmu sú v poslednom čase stále aktuálnejšie (Segond - Jensen, 1992; Rosenberg, 1985; Capelli - Ferrari, 1985).

Neuronálne siete majú oproti klasickému programovaniu určité principiálne prednosti. Prirodzeným spôsobom pracujú s neúplnou či neurčitou informáciou a preukazujú určité schopnosti pri samoučení ako aj samoodučení poznávacích vzorcov. Oblasť prirodzeného jazyka je svojím „rozmazaným" (fuzzy) charakterom pre aplikáciu konekčných modelov vhodná.

Za vynikajúcu myšlienku pokladáme pokus aplikovať na analýzu prirodzených jazykov *Boltzmanove stroje*. Boltzmanov stroj je fyzikálny systém, ktorý je najprv rozžeravený na vysokú teplotu a miera pohybu a neusporiadanosti jeho molekúl je vysoká. Počas poklesu teploty si jednotlivé jeho prvky hľadajú minimálny energetický stav, pričom vzniká usporiadanie. Ide teda o druh *metódy simulovaného žíhania*.

Môžeme si predstaviť napríklad zahriaty roztok, v ktorom sa počas chladnutia utvárajú kryštály. Alebo kalenie ocele: oceľ sa rozžeraví a dodáva sa jej energia (údermi kováčskeho kladiva). Počas chladnutia si molekuly kovu nachádzajú minimálny energetický stav, ktorý je zároveň aj najpevnejší.

Pri jazykovej analýze to znamená, že sa minimalizuje potenciálna energia (napätie) medzi vetnou formou a významom. Lokálne miminum sa pritom hľadá vzhľadom na všetky faktory súčasne (minimalizuje sa naraz syntaktické, sémantické aj kontextové napätie). Tento prístup je analógiou nášho multiagentového modelu SAPFO, hoci SAPFO nie je implementované konekcionisticky, ale diskrétne.

0.13 Všeobecnosť versus hĺbka

Medzi horizontálnym a vertikálnym rozmerom prirodzenojazykových systémov nastáva kompromis. *Čím je systém všeobecnejší, tým menej môže ísť do hĺbky.*

V systémoch spĺňajúcich *podmienku uzavretého sveta* (closed-world assumption) sa dá zadefinovať *konečno-doménová sémantika* (finite-domain semantics and reasoning, Palmer, 1990). V takých systémoch je lexika ohraničená, jej kolokačno-kombinačné možnosti, ako aj polysémia sa výrazne redukujú. Najneskôr pri inferencii možno každý jazykovo naznačený vzťah jednoznačne priradiť k niektorému zo sémantických pádov (prípadov, prototypov situácií), ktoré v takomto konečnom svete prichádzajú do úvahy a ktorých počet je tiež silno konečný.

Príkladmi takýchto systémov sú:

SHRDLU (Winograd, 1972). Rečové ovládanie robota vo svete kociek.

LUNAR (Woods, 1977). Systém asistujúci pri analýze mesačných vzoriek.

GUS (Bobrow, 1977). Systém na objednávanie leteniek.

LIFER (Hendrix, 1978). Databázový systém amerického námorníctva.

SOPHIE (Brown - Burton, 1975). Systém asistujúci pri návrhu integrovaných obvodov.

UNIX CONSULTANT (Wilensky, 1988) Sprievodca operačným systémom Unix.

METEO (Kittredge, 1973) Automatický anglicko-francúzsky preklad správ o počasí v Quebecu.

PORT (Tseytin, 1986) Organizácia prístavu.

Aplikované systémy sú teda úspešné vďaka svojej špecializácii. A všeobecné systémy sa nedarí aplikovať. Systém s neohraničeným diskurzom by musel byť schopný pracovať s nevyčerpateľným množstvom významových odtienkov a predpokladať ľubovoľné možné komunikatívne zámery. Ak ale zmyslová realita má mohutnosť kontinua, neexistuje funkcia, ktorá by ju jedno-jednoznačne zobrazovala na nejaký konečný systém. Možná je iba faktorizácia významového kontinua na konečný počet v sebe nekonečných tried. Jeden sémantický pád alebo iný významový komponent potom zahŕňa celú triedu javov, ktoré v rámci neho splývajú, sú nerozlíšiteľné.

Všeobecný model sa dá budovať iba iteratívne: postupným a nikdy sa nekončiacim zjemňovaním sémantiky podľa *princípu kontrastívnej diferenciácie* (0.16). Senzitivita systému je daná počtom sémantických pádov, sémantických dištinktívnych príznakov a komponentov významu (5.3, 1.3, 11.3).

Zistíme napríklad, že počítač generuje parafrázy:

bicyklovať
jazdiť na bicykli
**jazdiť v bicykli.*

Aby sme dosiahli rozlíšenie *‚jazdiť v aute'*, ale nie *‚jazdiť v bicykli'*, musíme zaviesť dištinktívny príznak ±*dutosť*. Dovtedy *auto* a *bicykel* splývajú ako nerozlíšiteľné v rámci triedy [+ins, +artef].

SAPFO je budované ako *všeobecný* systém. To znamená, že významová doména je neohraničená a za korektnosť parafráz ručíme iba obmedzene. Dôvodom toho je, že SAPFO má slúžiť v prvom rade ako nástroj na experiment so vzťahom kognitívno-percepčných štruktúr a jazyka. Cieľom je najprv získať poznatky o univerzáliách na rozhraní jazykového významu a obsahu. Tieto poznatky potom možno využiť pri tvorbe špecializovaných aplikovaných systémov.

Kritériom funkčnosti takýchto systémov potom môže byť napríklad schopnosť správne zodpovedať otázky na základe informácie v databanke alebo schopnosť zareagovať na výzvu správnym konaním.

Schopnosť parafrázovať môže slúžiť ako funkčné kritérium len pri podmienke všeobecnosti. Špecializovaný systém môže v dialógu obstáť, ale je tu nebezpečie, že sa tak stalo sčasti iba vďaka dômyselnému systému trikov. Ak by sme teda zvolili

prácu s ohraničeným modelom hneď na začiatku, výsledok by mohol byť efektnejší, ale získané poznatky by mohli stratiť všeobecnú kognitívnu hodnotu.

0.14 Experiment Par Ex (Paraphrasing of Expressions)

Za základ systému SAPFO slúžil nasledujúci psycholingvistický experiment:

Pokusnej osobe sa prečíta krátky príbeh (časť rozprávky). Potom sa vyzve, aby príbeh zopakovala. Príbehy musia byť zmysluplné a logicky uzavreté. Každý príbeh obsahuje okolo 100 slov, čiže je dostatočne krátky na to, aby si pokusná osoba zapamätala obsah, a zároveň dostatočne dlhý na to, aby nedokázala mechanicky zopakovať doslovné znenie.

Prerozprávania príbehu sa zaznamenajú a konfrontujú s pôvodnou verziou. Zatiaľ čo pokusné osoby sa domnievajú, že verne reprodukujú to, čo počuli, prerozprávané verzie vykazujú značné odchýlky. Pri reprodukcii pokusné osoby pozmeňujú lexiku, používajú rôznorodé syntaktické konštrukcie, dokonca pripájajú tvrdenia, ktoré sa v pôvodnom príbehu bezprostredne nenachádzali.

Príčinou týchto zmien je, že myseľ nezaznamenáva príbeh slovo po slove (ako reťazec znakov), ale pamätá si iba „esenciu obsahu", spolu s pocitmi a obrazmi, ktoré sa jej vynárali pri počúvaní príbehu. To je potom to, čo pokusné osoby neskôr opakujú.

Predpokladáme teda, že v jazykovom vedomí existuje nejaká *hĺbková reprezentácia* alebo *banka dát* (Fillmore, 1966; Horecký, 1991), hoci niektorí autori pripúšťajú, že by táto reprezentácia nemusela mať explicitnú formu (Winograd - Flores, 1987). Myseľ teda transformuje lineárnu povrchovú štruktúru jazyka na hĺbkovú reprezentáciu. Nová informácia zároveň efektívne splynie s informáciou, ktorá je už v banke dát prítomná. Neskôr potom myseľ nanovo generuje povrchové vetné formy zodpovedajúce hĺbkovej reprezentácii.

Tu však existuje množstvo rozličných ciest, ako externalizovať vnútornú reprezentáciu poznatkov. Voľba vonkajšej formy závisí od celého radu faktorov, ako sú kontext, konsituácia, aktuálne členenie, perspektíva, štýl, citový postoj a presvedčenia hovoriaceho, úroveň chápania a informovanosti adresáta a mnoho ďalších. Tak vznikajú parafrázy.

Takto môžeme študovať charakter hĺbkovej reprezentácie významu a vzťah medzi *hĺbkovým* a *povrchovým* pomocou **porovnávania všetkých povrchových foriem, ktoré hovoriaci asociuje s jedným a tým istým obsahom.**

PRINCÍPY

0.15 Princíp jedno-jednoznačnosti reprezentácie

Priradenie medzi hĺbkovou reprezentáciou a významom musí byť jedno-jednoznačné (bijektívne);

frázam s rovnakým (blízkym) významom musí zodpovedať rovnaká (blízka) reprezentácia; frázam s rozdielnym významom musí zodpovedať rozdielna reprezentácia.

Formálne: ak máme dve frázy f_1, f_2 s významom $V(f_1)$, $V(f_2)$ a ďalej konštruktívne dané zobrazenie

R: povrchové štruktúry -> hĺbkové štruktúry

a vzťah *reprezentácie* ≡, potom ak $R(f_i)$ reprezentuje $V(f_i)$:

$$R(f_1) \equiv V(f_1)$$
$$R(f_2) \equiv V(f_2)$$

musí aj

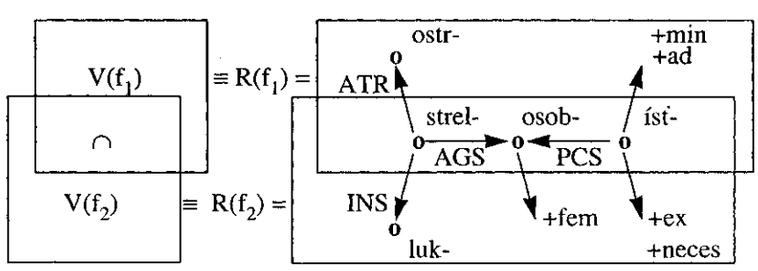
$$R(f_1) \cap R(f_2) \equiv V(f_1) \cap V(f_2)$$
$$R(f_1) - R(f_2) \equiv V(f_1) - V(f_2)$$
$$R(f_2) - R(f_1) \equiv V(f_2) - V(f_1)$$

teda prienik reprezentácií musí reprezentovať prienik významov a rozdiel reprezentácií musí reprezentovať rozdiel významov.

Čiže reprezentácia musí byť karteziánskym súčinom komponentov významu, v ktorom každý komponent významu je explicitne viditeľný a nezávisle manipulovateľný.

Príklad.

f_1 = Ostrostrelec neprišiel.
f_2 = Lukostrelkyňa musí odísť.



0.16 Princíp kontrastívnej diferenciácie

Princíp vydeľovania sémantických pádov, dištinktívnych príznakov a komponentov významu na základe porovnávania povrchovej formy významovo usúvzťažnených konštrukcií.

1. Diferenciácia sémantických pádov. E. Tibenská (1989) študovala konštrukcie typu:

Družstevníci naplnili voz obilím.^{REA} → Obilie naplnilo voz.
Družstevníci naplnili voz lopatami.^{INS} *→ Lopaty naplnili voz.

Prvé dve konštrukcie sú usúvzťažnené, druhé dve nemusia byť (ak nejde naozaj o voz naplnený lopatami). Z rozdielnosti povrchových štruktúr vyvodila existenciu nového hĺbkového pádu a dovtedajší INŠTRUMENTÁL rozčlenila na vlastný sémantický INŠTRUMENTÁL a *REALIZÁTOR*, t. j. entity, ktoré nefungujú ako nástroje, ale realizujú dej svojou vlastnou substanciou. Keďže sa tieto dva pády prejavujú rozličným súborom výrazových prostriedkov, je potrebné rozlíšiť ich aj v hĺbkovej reprezentácii. Ak by v hĺbkovej reprezentácii splývali, pomiešali by sme ich aj pri generovaní povrchových štruktúr.

Iný príklad. Vo výrazoch *byť na stole, pod stolom, pred stolom, za stolom* ... ide napospol o pád LOKATÍV, nie sú však vzájomne ľubovoľne zameniteľné:

pri stole → vedľa stola, v stole *→ vedľa stola

Správne transformácie si vynucujú zaviesť jemnejšie sekundárne rozlíšenie LOKATÍVu, ktoré budeme označovať arabskými číslicami: LOK_1, LOK_2, LOK_3...

2. Diferenciácia sémantických dištinktívnych príznakov. Príklad.

Odrezal som Janovi^{BEN} chlieb. → Odrezal som chlieb pre Jana.
Odrezal som Janovi^{PAR} ruku. → Odrezal som Janovu ruku.

Interpretácia datívu v týchto konštrukciách závisí od sémantickej triedy objektu v akuzatíve. V prvom prípade je Jano v role BENEFAKTORa, v druhom je ruka jeho fyzickou časťou. Odlíšenie druhého typu konštrukcií teda vyžaduje zavedenie sémantickej triedy *korp* (časti tela).

Iný príklad.

piecť múku^{MAT} → piecť múku na chlieb, piecť chlieb z múky
piecť chlieb^{REZ} *→ piecť chlieb na múku, piecť múku z chleba

Schopnosť generovať správne transformácie vyžaduje rozlišovať medzi MATERIÁLom a REZULTANTom *pečenia*; a schopnosť rozlišovať ich ďalej vyžaduje zaviesť sémantický dištinktívny príznak +*artef* (artefakty).

Iný príklad.

jazdiť autom → jazdiť na aute → jazdiť v aute
jazdiť bicyklom → jazdiť na bicykli *→ jazdiť v bicykli

Systém SAPFO generuje parafrázu *jazdiť na bicykli*, ale aj nesprávnu parafrázu *jazdiť v bicykli*. Na základe valenčných pravidiel totiž inštrumenty pohybových slovies ako také splývajú.

VAL(pohyb.sloveso, [ø+inš;na+lok;v+lok],[+instr]) = INS

Ak sa chceme vyhnúť nesprávnemu použitiu predložkového pádu *v+lokál*, máme možnosť rozšíriť systém dištinktívnych príznakov o príznak *+dut* (dutosť). Potom môžeme valenčné rovnice prepísať:

VAL(pohyb.sloveso, [ø+inš;na+lok],[+instr]) = INS
VAL(pohyb.sloveso, [v+lok],[+instr,+dut]) = INS

Všeobecne: Majme konštrukciu 1 s lexikálnym obsadením charakterizovaným súborom príznakov $+\delta_1...+\delta_n$, ktorá sa dá transformovať na konštrukciu 2. Ak existuje iné konkrétne lexikálne obsadenie, ktoré takúto transformáciu neumožňuje napriek tomu, že je charakterizované rovnakou n-ticou $+\delta_1...+\delta_n$:

$$\text{konštrukcia}_1(+\delta_1...+\delta_n) \rightarrow \text{konštrukcia}_2$$
$$\text{konštrukcia}_1(+\delta_1...+\delta_n) \rightarrow {}^*\text{konštrukcia}_2$$

treba zaviesť dištinktívny príznak $+\delta_{n+1}$ tak, aby prípustnosť či neprípustnosť danej transformácie rozlišoval:

$$\text{konštrukcia}_1(+\delta_1...+\delta_n,+\delta_{n+1}) \rightarrow \text{konštrukcia}_2$$
$$\text{konštrukcia}_1(+\delta_1...+\delta_n,-\delta_{n+1}) \rightarrow \emptyset$$

3. Diferenciácia komponentov významu. Príklad.

Porovnajme adjektiváliá *žltý/zožltnutý*. Táto dvojica je očividne významovo usúvzťažnená. *Zožltnutý* sa významovo zhoduje so *žltý*, plus je chápané ako výsledok dynamického procesu. Táto sémantická diferencia je elementárna a v slovenčine funguje systematicky. Budeme ju nazývať *rezultatívnosť* (Sejáková, 1988) a písať:

zožltnutý = žltý +rez

Všeobecne: ak máme dve usúvzťažnené konštrukcie:

$$\text{konštrukcia}_1 \approx \text{konštrukcia}_2$$

pričom ich významy

$$V(\text{konštrukcia}_1) = V(\text{konštrukcia}_2) + \delta$$

potom ak sémantická diferencia δ je elementárna (ďalej nerozložiteľná) a v danom jazyku funguje systematicky, nazveme ju *komponentom významu*.

0.17 Synergický princíp

Princíp spracovania jazyka paralelnými a neobmedzene komunikujúcimi modulmi. Vzniká tým synergický efekt, t. j. schopnosť systému podať výkon, ktorý prevyšuje sumu výkonov jeho jednotlivých častí, ak by pracovali oddelene.

Synergiu sme vysvetlili v odseku 0.11. Súčinnosť a synergický efekt sa uplatňujú pri morfologickej, syntaktickej, sémantickej, ako aj slovotvornej analýze.

0.18 Príncíp 51 %

Do systému zaraďujeme všetky pravidlá, ktoré fungujú aspoň v 51 %-tách prípadov.

Pretože naším cieľom je minimalizovať celkovú chybu (0.5), t. j. minimalizovať počet nesprávne vygenerovaných + nevygenerovaných ale správnych parafráz, vyplýva z toho, že môžeme a musíme využiť každé konštrukčné pravidlo, ktoré funguje s viac ako 50 %-nou pravdepodobnosťou. Zaradiť takéto pravidlo znamená, že v 49 prípadoch zo 100 nastane chyba 1. druhu. Ale nezaradiť takéto pravidlo znamená, že v 51 prípadoch zo 100 nastane chyba 2. druhu. Nezaradením pravidla sa teda celková chyba zväčší.

Navyše, vďaka synergickému efektu sa môže stať, že súčinnosťou viacerých 51 %-ných pravidiel, ktoré na seba vzájomne pôsobia selekčným efektom, vzniká výsledok s 80-90 %-nou pravdepodobnosťou korektnosti.

Už pri morfologickej analýze vstupujú do činnosti alternatívne pravidlá, ktorých pravdepodobnosť býva aj menšia ako 50 %. Ale vďaka selekčnému efektu syntaxe a kontextu je výslednicou použitia 20-30 %-ných pravidiel 99-100 %-ný výsledok. To isté platí aj pre sémantické a slovotvorné konštrukčné pravidlá.

Poznámka. Zrejme nie je problém vyrobiť program, ktorý nerobí žiadne chyby 2. druhu, ale zato veľa chýb 1. druhu. Je to tento:

```
for I1=0 to 256
    for I2=0 to 256
        ...
            for I100=0 to 256
                print string[I1,I2,...,I100]
                new line
            next I100
        ...
    next I2
next I1
```

Nie je ťažké vyrobiť ani program, ktorý nerobí žiadne chyby 1. druhu, ale zato veľa chýb 2.druhu. Napríklad:

```
begin
    rem no command
end
```

Problém je minimalizovať chyby 1. a 2. druhu *súčasne*.

1.

SLOVNÍKY

1.1 Údajová a procedurálna časť

Čo do formy, celý systém programov SAPFO je tvorený jednotným formalizmom. Pozostáva z n-árnych logických predikátov typu

$$\text{predikát}(arg_1, arg_2, arg_3, \ldots, arg_n),$$

pospájaných logickými spojkami konjunkcie, disjunkcie, implikácie a symbolom rezu.

Čo do obsahu, program sa delí na časť údajovú a procedurálnu. Údajovú časť tvoria slovníky: morfologický, lexikálno-sémantický, slovotvorný, syntaktickoväzobný, valenčný, synonymický, frazeologický, štylistický a konotačný slovník. Procedurálnu časť tvoria algoritmy morfologickej, morfematickej, syntaktickej, sémantickej a kontextovej analýzy a syntézy.

Slovníky obsahujú minimálnu nevyhnutnú informáciu, ktorá nie je vyvoditeľná zo všeobecných zákonitostí jazyka, ale je individuálna pre každú lexému. Je to teda určitá minimálna miera údajov, na ktorú počítač nemôže prísť sám ani ju vyvodiť dedukciou zo všeobecnejších pravidiel, ale musí byť poskytnutá manuálne človekom.

Procedurálna časť obsahuje naproti tomu algoritmy a pravidlá všeobecného charakteru, platné pre celú triedu lexikálnych jednotiek, charakterizovanú určitou všeobecnou jazykovou kategóriou alebo subkategoriálnym príznakom. Stelesňuje všeobecné zákonitosti jazyka, jeho vnútornú logiku.

Oddelenie slovníkovej a neslovníkovej informácie sprehľadňuje celý jazykový systém a zjednodušuje jeho formalizáciu.

1.2 Morfologický slovník

Morfologický slovník obsahuje ku každej ohybnej lexikálnej jednotke všetky tie údaje, ktoré sú potrebné na jej gramatické ohýbanie. Podľa jednotlivých slovných druhov sú to:

substantíva:	koreň,
	rod,
	počítateľnosť,
	skloňovací vzor s alternáciou
adjektíva:	koreň,
	skloňovací vzor,
	stupňovací vzor s alternáciou

slovesá:	infinitívny kmeň,
	prézentný kmeň,
	zvratnosť,
	vid,
	časovací vzor s alternáciou
príslovky:	koreň,
	prípona pozitívu,
	stupňovací vzor
predložky:	základný tvar,
	vokalizovaný tvar
spojky:	základný tvar
častice:	základný tvar
citoslovcia:	základný tvar

Príklad. Časť morfologického slovníka

```
substantívum(záhradník,muž,počítateľné,vodník).
substantívum(líst,str,hromadné,vysvedčenie).
substantívum(nožnic,žen,pomnožné,ulica).
adjektívum(Aničkin,matkin,nil).
adjektívum(čarovn,pekný,belasý).
adjektívum(krásn,krásny,supletívne).
sloveso(voň,von,nezvratné,nedokonavé,klaňať).
sloveso(rozhor,rozhor,sa,dokonavé,vidieť).
sloveso(darov,daruj,si,obojvidové,pracovať).
príslovka(bos,o,nil).
príslovka(pekn,e,supletívne).
príslovka(perzsk,y,belasý).
predložka(s,so).
predložka(k,ku).
predložka(pomocou,pomocou).
spojka(a).
spojka(ale).
spojka(či).
častica(azda).
častica(žiaľ).
častica(zrejme).
citoslovce(ach).
citoslovce(mňau).
citoslovce(žblnk).
```

Skloňovací vzor substantíva

je definovaný dvanásticou pádových prípon a zoznamom alternácií.

Podľa toho *mak* je iný vzor ako *dub*, pretože sa líšia v lokáli singuláru (*o dub<u>e</u>, o mak<u>u</u>*). *Perla* a *čarodejka* majú síce rovnakú paradigmu prípon, ale líšia sa alternáciou (*od per<u>á</u>l, od čarodej<u>o</u>k*). Ide teda tiež o odlišné vzory.

Spracovaním materiálu v Morfológii slovenského jazyka (1966) sme dostali 132 skloňovacích vzorov pre slovenské podstatné mená. Len vzor dub sa rozpadne na 22 podvzorov. Voľba jedného z týchto 132 vzorov dáva jednoznačnú informáciu pre skloňovanie podstatného mena za predpokladu, že poznáme jeho koreň a počítateľnosť.

Počítateľnosť

Počítateľnosť podstatného mena je z trojprvkovej množiny = {počítateľné, hromadné, pomnožné}.

Ak zohľadníme počítateľnosť, dostaneme ďalší systém podvzorov. Prekrývajú sa vždy s nejakým hlavným vzorom, ale s tou podmienkou, že z dvanástice pádových prípon je prípustná len šestica singulárových (pri hromadných menách) alebo plurálových (pri pomnožných menách).

Rod

Rod podstatného mena je z trojprvkovej množiny = {ženský, mužský, stredný}. Nepriamo vyplýva zo skloňovacieho vzoru, takže ho ani netreba uvádzať samostatne.

Časovací vzor slovesa

je daný sedmicou tvarových prípon v týchto siedmich tvaroch: infinitív, l-príčastie, v-príčastie, n/t-príčastie, prézentný kmeň, imperatív, prechodník. Týchto sedem tvarov je kľúčových. Neurčitkový a prítomníkový kmeň spolu so sedmicou tvarových prípon predstavujú minimálnu a postačujúcu informáciu pre tvorenie všetkých ostatných (až 247) slovesných tvarov. Vychádza 24 jednoznačne definovaných slovesných vzorov.

Zvratnosť slovesa

Reflexivita slovies je z trojprvkovej množiny {#,sa,si}.

Vid slovesa

Slovesný vid je z trojprvkovej množiny {dokonavé, nedokonavé, obojvidové}.

Skloňovací vzor adjektíva

Adjektívny skloňovací vzor je daný 36-ticou pádovo-rodových prípon. Postačujúcou informáciou pre skloňovanie prídavného mena je jeho koreň a 36-tica prípon. Jednoznačne definovaných adjektívnych vzorov je 7.

Stupňovací vzor adjektíva (príslovky)

Existuje 6 vzorov, ktoré sa líšia alternáciou, a supletívne stupňovanie. Postačujúcu informáciu tvoria koreň a stupňovací vzor.

Podľa tých istých vzorov sa stupňujú aj stupňovateľné adverbiá.

Zámená a číslovky

preberajú gramatické kategórie tých slovných druhov, v ktorých funkcii sa nachádzajú. Vyčerpávajúci deklaratívny popis ich morfológie sme už uvideli (Páleš, 1989).

Skloňovacie vzory substantív		
vzor	*paradigma pádových prípon*	*alternácia*
chlap	ø,a,ovi,a,ovi,om,i,ov,om,ov,och,mi	
peržan	ø,a,ovi,a,ovi,om,ia,ov,om,ov,och,mi	
dedo	o,a,ovi,a,ovi,om,ovia,ov,om,ov,och,ami	
feláh	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,mi	
botto	o,u,ovi,u,ovi,om,ovia,ov,om,ov,och,ami	
hosť	ø,a,ovi,a,ovi,om,ia,í,om,í,och,ami	*
boh	ø,a,u,a,u,om,ovia,ov,om,ov,och,mi	
mím	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,ami	
lovec	ø,a,ovi,a,ovi,om,i,ov,om,ov,och,ami	»« -1sg
vodník	ø,a,ovi,a,ovi,om,i,ov,om,ov,och,mi	k/c +1pl
černoch	ø,a,ovi,a,ovi,om,i,ov,om,ov,och,mi	ch/s +1pl
hrdina	a,u,ovi,u,ovi,om,ovia,ov,om,ov,och,ami	
otec	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,ami	»« -1sg
turek	ø,a,ovi,a,ovi,om,i,ov,om,ov,och,ami	k/c +1pl
		»« -1pl
paris	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,mi	s/d -1sg
aias	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,mi	s/n -1sg
nero	ø,a,ovi,a,ovi,om,i,ov,om,ov,och,mi	+n, ≤ -1sg
anakreón	ø,a,ovi,a,ovi,om,i,ov,om,ov,och,mi	+t, ≥ -1sg
achilles	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,mi	-es -1sg
syzifos	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,mi	-os -1sg
anaxagoras	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,mi	-as -1sg
génius	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,mi	-us -1sg
pontifex	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,mi	ex/ik -1sg
noe	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,ami	+m -1sg
futbalista	a,u,ovi,u,ovi,om,i,ov,om,ov,och,ami	
bandita	a,u,ovi,u,ovi,om,i,ov,om,ov,och,mi	
hume	ø,a,ovi,a,ovi,om,ovia,ov,om,ov,och,ami	
vrchný	ý,ého,ému,ého,om,ým,í,ých,ým,ých,ých,ými	
cestujúci	i,eho,emu,eho,om,im,i,ich,im,ich,ich,imi	
dub	ø,a,u,ø,e,om,y,ov,om,y,och,mi	
mak	ø,u,u,ø,u,om,y,ov,om,y,och,mi	
med	ø,u,u,ø,e,om,y,ov,om,y,och,mi	
svietnik	ø,a,u,ø,u,om,y,ov,om,y,och,mi	
papier	ø,a,u,ø,i,om,e,ov,om,e,och,mi	
model	ø,u,u,ø,i,om,y,ov,om,y,och,mi	
strom	ø,u,u,ø,e,om,y,ov,om,y,och,ami	
meter	ø,a,u,ø,i,om,e,ov,om,e,och,ami	»« -14sg
hotel	ø,a,u,ø,i,om,y,ov,om,y,och,mi	
rytmus	ø,u,u,ø,e,om,y,ov,om,y,och,ami	-us -14sg

vzor	*paradigma pádových prípon*	*alternácia*
nónius	ø,a,u,ø,u,om,y,ov,om,y,och,ami	-us -14sg
most	ø,a,u,ø,e,om,y,ov,om,y,och,ami	
zlomok	ø,u,u,ø,u,om,y,ov,om,y,och,ami	»« -14sg
šperk	ø,u,u,ø,u,om,y,ov,om,y,och,ami	
ohníček	ø,a,u,ø,u,om,y,ov,om,y,och,ami	»« -14sg
pojem	ø,u,u,ø,e,om,y,ov,om,y,och,ami	»« -14sg
papier	ø,a,u,ø,i,om,e,ov,om,e,och,mi	
chrbát	ø,a,u,ø,e,om,y,ov,om,y,och,mi	»« -14sg
chlieb	ø,a,u,ø,e,om,y,ov,om,y,och,mi	≥ -14sg
mráz	ø,u,u,ø,e,om,y,ov,om,y,och,mi	≥ -14sg
stôl	ø,a,u,ø,e,om,y,ov,om,y,och,mi	≥ -14sg
vietor	ø,a,u,ø,e,om,y,ov,om,y,och,mi	≥,»«-14sg
stroj	ø,a,u,ø,i,om,e,ov,om,e,och,mi	
jeleň	ø,a,ovi,a,ovi,om,e,ov,om,e,och,mi	
mravec	ø,a,ovi,a,ovi,om,e,ov,om,e,och,ami	»« -1sg
čaj	ø,u,u,ø,i,om,e,ov,om,e,och,mi	
veniec	ø,a,u,ø,i,om,e,ov,om,e,och,ami	»« -14sg
dážď	ø,a,u,ø,i,om,e,ov,om,e,och,ami	≥ -14sg
had	ø,a,ovi,a,ovi,om,y,ov,om,y,och,mi	
orol	ø,a,ovi,a,ovi,om,y,ov,om,y,och,ami	
peniaz	ø,u,u,ø,i,om,e,í,om,e,och,mi	≥ +2pl
plášť	ø,a,u,ø,i,om,e,ov,om,e,och,ami	
deň	ø,a,u,ø,i,om,i,í,om,i,och,ami	»« -14sg
kôň	ø,a,ovi,a,ovi,om,e,í,om,e,och,mi	≥ -1sg
kuli	ø,ho,mu,ho,m,m,ovia,ov,om,ov,och,ami	
atašé	ø,ø,ø,ø,ø,ø,ø,ø,ø,ø,ø,ø,ø	
žena	a,y,e,u,e,ou,y,ø,ám,y,ách,ami	≤ +2pl
zora	a,y,e,u,e,ou,e,ø,ám,y,ách,ami	≤ +2pl
matka	a,y,e,u,e,ou,y,ø,ám,y,ách,ami	«ie» +2pl
vojna	a,y,e,u,e,ou,y,ø,ám,y,ách,ami	«e» +2pl
jamka	a,y,e,u,e,ou,y,ø,ám,y,ách,ami	«ô» +2pl
perla	a,y,e,u,e,ou,y,ø,ám,y,ách,ami	«á» +2pl
kráska	a,y,e,u,e,ou,y,ø,am,y,ach,ami	«o» +2pl
čarodejka	a,y,e,u,e,ou,y,ø,ám,y,ách,ami	«o» +2pl
idea	a,y,e,u,i,ou,y,í,ám,y,ách,ami	
dáma	a,y,e,u,e,ou,y,ø,am,y,ach,ami	
úloha	a,y,e,u,e,ou,y,ø,ám,y,ách,ami	
izis	ø,y,e,u,e,ou,y,ø,ám,y,ách,ami	s/d -1sg
ceres	ø,y,e,u,e,ou,y,ø,ám,y,ách,ami	s/r -1sg
juno	o,y,e,u,e,ou,y,ø,am,y,ach,ami	+n -1sg
demeter	ø,y,e,u,e,ou,y,ø,ám,y,ách,ami	»«-1sg-2pl ≤ +2pl

vzor	*paradigma pádových prípon*	*alternácia*
víchrica	a,e,i,u,i,ou,e,ø,iam,e,iach,ami	
ulica	a,e,i,u,i,ou,e,ø,iam,e,iach,ami	≤ +2pl
funkcia	a,e,i,u,i,ou,e,í,ám,e,ách,ami	
sudkyňa	a,e,i,u,i,ou,e,ø,ám,e,iach,ami	≤ +2pl
vládkyňa	a,e,i,u,i,ou,e,ø,ám,e,iach,ami	
čerešňa	a,e,i,u,i,ou,e,ø,iam,e,iach,ami	«ie» +2pl
ruža	a,e,i,u,i,ou,e,í,iam,e,iach,ami	
vôňa	a,e,i,u,i,ou,e,í,am,e,ach,ami	
svieca	a,e,i,u,i,ou,e,ø,am,e,ach,ami	
dlaň	ø,e,i,ø,i,ou,e,í,iam,e,iach,ami	
myseľ	ø,e,i,ø,i,ou,e,í,iam,e,iach,ami	»« -14sg
báseň	ø,e,i,ø,i,ou,e,í,am,e,ach,ami	»« -14sg
koľaj	ø,e,i,ø,i,ou,e,í,am,e,ach,ami	
pani	i,ej,ej,iu,ej,ou,ie,í,iam,ie,iach,iami	
kráľovná	á,ej,ej,ú,ej,ou,é,ø,ám,é,ách,ami	«ie» +2pl
gazdiná	á,ej,ej,ú,ej,ou,é,ø,ám,é,ách,ami	≤ +2pl
ženská	á,ej,ej,ú,ej,ou,é,ých,ým,é,ých,ými	
mať	ø,e,i,ø,i,ou,e,í,iam,e,iach,ami	+er -1sg
kosť	ø,i,i,ø,i,ou,i,í,iam,i,iach,ami	
cirkev	ø,i,i,ø,i,ou,i,í,iam,i,ách,ami	»« -14sg
kader	ø,e,i,ø,i,ou,e,í,ám,e,ách,ami	
labuť	ø,i,i,ø,i,ou,e,í,iam,e,iach,ami	
sapfo	ø,ø,ø,ø,ø,ø,ø,ø,ø,ø,ø,ø,ø	
mesto	o,a,u,o,e,om,'á',ø,ám,á,ách,ami	≤ +2pl
mestečko	o,a,u,o,u,om,á,ø,ám,á,ách,ami	«ie» +2pl
rádio	o,a,u,o,u,om,á,í,ám,á,ách,ami	
jedlo	o,a,u,o,u,om,á,ø,ám,á,ách,ami	
sklenárstvo	o,a,u,o,e,om,a,ø,am,a,ach,ami	«ie» +2pl
slnko	o,a,u,o,u,om,á,ø,ám,á,ách,ami	≤ +2pl
krídlo	o,a,u,o,u,om,a,ø,am,a,ach,ami	«e» +2pl
rúcho	o,a,u,o,u,om,a,ø,am,a,ach,ami	
miesto	o,a,u,o,e,om,a,ø,am,a,ach,ami	
lýtko	o,a,u,o,u,om,a,ø,am,a,ach,ami	«o» +2pl
gesto	o,a,u,o,e,om,'á',ø,ám,á,ách,ami	
capriccio	io,a,u,o,u,om,á,í,ám,á,ách,ami	
fórum	ø,a,u,ø,e,om,a,ø,am,a,ach,ami	-um -14sg
múzeum	ø,a,u,ø,u,om,á,í,ám,á,ách,ami	um -14sg
srdce	e,a,u,e,i,om,ia,ø,iam,ia,iach,ami	≤ +2pl
riečište	e,a,u,e,i,om,ia,ø,iam,ia,iach,ami	
drievce	e,a,u,e,i,om,ia,ø,iam,ia,iach,ami	«e» +2pl
líce	e,a,u,e,i,om,a,ø,am,a,ach,ami	
citoslovce	e,a,u,e,i,om,ia,ø,iam,ia,iach,ami	«ie» +2pl

vzor	*paradigma pádových prípon*	*alternácia*
more	e,a,u,e,i,om,ia,í,iam,ia,iach,iami	
oje	e,a,u,e,i,om,e,í,am,e,ach,ami	
vysvedčenie	ie,ia,iu,ie,í,ím,ia,í,iam,ia,iach,iami	
dievča	a,a,u,a,i,om,á,ø,ám,á,ách,ami	+ať -14sg
mača	a,a,u,a,i,om,á,ø,ám,á,ách,ami	+ať -14sg ≤ +2pl
kura	a,a,u,a,i,om,e,ø,om,e,och,ami	+ať+2367sg +enc +pl
teľa	a,a,u,a,i,om,e,ø,om,e,och,ami	+ať+2367sg +c pl
dieťa	a,a,u,a,i,om,i,í,om,i,och,mi	+ať+2367sg ≥ pl
cestovné	é,ého,ému,é,om,ým,é,ých,ým,é,ých,ými	

Substantívne koreňové alternácie			
		Vokalické	
≤	predĺženie samohlásky	a/á/ia, i/í, y/ý, u/ú o/ô e/é/ie, r/ŕ, l/ĺ, ä/ia	vlna/vĺn, voda/vôd žena/žien, laba/láb
≥	skrátenie samohlásky	á/a, í/i, ú/u, ie/e, ia/a ô/o	stôl/stola, chlieb/ chleba, mráz/mrazu
»«	vypustenie samohlásky	e/ø, o/ø, i/ø, á/ø, ie/ø	dravec/dravca švagor/švagra
«»	vloženie samohlásky	ø/ie, ø/e, ø/o, ø/ô, ø/á	kráska/krások, perla/perál
		Konsonantické	
/	zámena spoluhlásky	k/c, s/d, s/r, s/n, ch/s	Turek/Turci Izis/Izidy
+	rozšírenie koreňa	+n, +m, +t, +ať, +enc, +c +ik, +er	dievča/dievčatá Noe/Noema
-	zúženie koreňa	-us, -os, -es, -as, -um, -ex	génius/génia Aiolos/Aiola
*	lingválna/ dorzálna	ľ/l, ť/t, ď/d	loď/lode kosť/kosti

Časovacie vzory slovies		
Vzor	*Sedmica kľúčových morfém*	*Zoznam alternácií*
chytať	a,a,av,an,á,aj,ajú	
klaňať	a,a,av,an,ia,aj,ajú	A_5
čítať	a,a,av,an,a,aj,ajú	
rozumieť	ie,e,ev,en,ie,ej,ejú	≥
niesť	ø,o,ø,en,ie,ø,ú	A_6
viesť	ø,o,ø,en,ie,ø,ú	A_2, A_5, A_6
hynúť	ú,u,uv,ut,ie,ø,ú	A_5
pracovať	a,a,av,an,e,ø,ú	
robiť	i,i,iv,en,í,ø,ia	
kúpiť	i,i,iv,en,i,ø,ia	
vidieť	ie,e,iv,en,í,ø,ia	A_5
kričať	a,a,av,an,í,ø,ia	
trieť	ie,e,ev,et,ie,i,ú	
brať	a,a,av,an,ie,ø,ú	A_5
česať	a,a,av,an,e,ø,ú	A_1, A_7
chudnúť	ú,o,uv,ut,e,i,ú	A_3
vládnuť	u,o,uv,ut,e,i,u	A_3
žuť	ø,ø,v,t,e,ø,ú	
kliať	ø,ø,v,t,e,ø,ú	
piť	ø,ø,v,t,e,ø,ú	A_4
žať	a,a,av,at,e,i,ú	
päť	ä,ä,äv,ät,e,i,ú	
kresliť	i,i,iv,en,í,i,ia	
krášliť	i,i,iv,en,i,i,ia	

Slovesné kmeňové alternácie			
A_1	inf.kmeň/préz.kmeň	t/c, d/dz, k/č, c/č, dz/dž, s/š, z/ž, ch/š, h/ž, sl/šľ	česať/češe plakať/plače blikotať/blikoce
A_2	inf.kmeň/préz.kmeň/ l-príčastie	s/d/ď, s/t/ť, c/č/k, c/ž/h, z/ž/h	tĺcť/tlčie/tĺkol húsť/hudie/húdol
A_3	Vypustenie derivačnej morfémy -n- v l-príčastí	n/ø	chudnúť/chudol vládnuť/vládol
A_4	Vypustenie interfigovanej morfémy -j- v imperatíve	j/ø	pije/pi
A_5	Zmena dorzálnej konsonanty na lingválnu	d/ď, t/ť, n/ň, l/ľ	hynúť/hyň
A_6	Tvorenie gerundia l- a t-príčastia od préz. kmeňa		
A_7	Tvorenie prechodníka od oboch kmeňov		

Skloňovacie vzory adjektiválií	
Vzor	*Skloňovacia paradigma*
pekný	ý,ého,ému,ého,om,ým
	ý,ého,ému,ý,om,ým
	á,ej,ej,ú,ej,ou).
	é,ého,ému,é,om,ým
	í,ých,ým,ých,ých,ými
	é,ých,ým,é,ých,ými
krásny	y,eho,emu,eho,om,ym
	y,eho,emu,y,om,ym
	a,ej,ej,u,ej,ou
	e,eho,emu,e,om,ym
	i,ych,ym,ych,ych,ymi
	e,ych,ym,e,ych,ymi
cudzí	í,ieho,iemu,ieho,om,ími
	í,ieho,iemu,í,om,ími
	ia,ej,ej,iu,ej,ou
	ie,ieho,iemu,ie,om,ím
	í,ích ,ím,ích,ích,ími
	ie,ích,ím,ie,ích,ími
rýdzi	i,eho,emu,eho,om,im
	i,eho,emu,i,om,im
	a,ej,ej,u,ej,ou
	e,eho,emu,e,om,im
	i,ich,im,ich,ich,imi
	e,ich,im,e,ich,imi
otcov	ø,ho,mu,ho,om,ým
	ø,ho,mu,ø,om,ým
	a,ej,ej,u,ej,ou
	o,ho,mu,o,om,ým
	i,ých,ým,ých,ých,ými.
	e,ých,ým,e,ých,ými
papagájí	í,eho,emu,eho,om,im
	í,eho,emu,í,om,im
	a,ej,ej,u,ej,ou
	e,eho,emu,e,om,im
	í,ích,ím,ich,ich,imi
	e,ích,ím,e,ich,imi
môj	ø,ho,mu,ho,om,ím
	ø,ho,mu,ø,om,ím
	a,ej,ej,u,ej,ou
	e,ho,mu,e,om,ím
	i,ich,im,ich,ich,imi
	e,ich,im,e,ich,imi

Stupňovacie vzory adjektív a adverbií	
Vzor	*Alternácie*
nový	
krátky	-k, ≥
vysoký	-ok, s/š, z/ž
ďaleký	-ek
biely	≥
belasý	+ej
supletívne	

1.3 Lexikálno-sémantický slovník

Nedostatočnosť syntaxe. Po prvých výskumoch v 50-tych rokoch (Chomsky, 1956), ktoré kládli dôraz hlavne na syntax, sa dostatočne zreteľne ukázalo, že samotná znalosť syntaxe na tvorenie viet nestačí. V rámci jednej syntaktickej kategórie sa slová líšia ešte lexikálnym významom, ktorý je taktiež rozhodujúci pre ich správanie sa vo vete:

Ivan beží.	*Ivan beží cestu.	povedal som, že...
*Ivan zahodil.	Ivan zahodil mincu.	*oholil som sa, že...
smutný, že...	povesť o Odyseovi	chuť pracovať
*červený, že...	*loď o Odyseovi	*dom pracovať

Je to časť významu, ktorá nie je na slove explicitne indikovaná žiadnymi viditeľnými príznakmi, ale dá sa zachytiť iba slovníkovo. V 70-tych a 80-tych rokoch sa potom venovalo veľké úsilie precíznemu rozpracovaniu lexikálnej sémantiky a tvorbe bohatých počítačových slovníkov a databánk.

Lexikálno-sémantický slovník SAPFO

obsahuje ku každej lexikálnej jednotke minimálnu lexikálno-sémantickú informáciu. Súbor sémantických príznakov zaraďuje každú lexému do určitých sémantických tried, ktoré sú pre jej správanie vo vete dištinktívne.

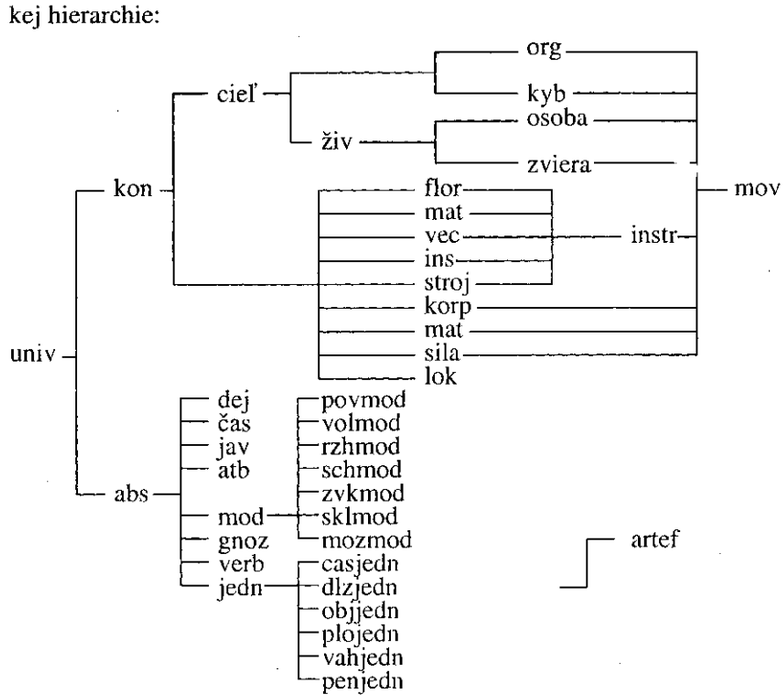
Príklad. Časť lexikálno-sémantického slovníka

chlieb	[+mat,+artef]
voda	[+mat;+sila]
dobrý	[+kval]
železný	[+klas]
mačací	[+drprivl]
padať	[2]
rozprávať	[60]
predstaviť si	[100]

doma	[lok$_1$]
domov	[des$_1$]
trochu	[mes]
dnes	[tmp$_1$]
nahlas	[mds]
okolo	[via$_6$]
napriek	[kcs]
s_ohľadom_na	[res$_1$]
a	[zlučov,simult]
tu	[zlučov,altern]
ale	[odpor,adverz]
a navyše	[stupň,grad]
lebo	[kaz]
ak	[knd]
keď	[tmp]
avšak	[uvádz,pripáj]
najmä	[vytyč,zdôraz]
zrejme	[hodnot,80%]
ajhľa	[vôľ,upozorň]
cink	[zvukmal,zvukov]

Dištinktívne príznaky substantív

Na základe princípu kontrastívnej diferenciácie (0.16) sme pre substantíva vyčlenili 42 dištinktívnych sémantických príznakov, usporiadaných do taxonomickej hierarchie:



Dedičnosť sémantických črt

Dištinktívne príznaky sú štruktúrované do inheritačnej hierarchie. Pri každom podstatnom mene stačí v slovníku uviesť iba minimálny súbor kardinálnych sémantických čŕt. Ostatné sú potom z hierarchie inferovateľné. Napríklad charakteristika *+osoba* implikuje príznaky *+životné, +cieľavedomé, +konkrétne, +hnuteľnosť*, ale aj *-inštrument* a *-abstraktné*.

Problém voľby sémantických tried

Ak by sme chceli použiť úplnejší systém tried, mohli by sme použiť rozdelenie objektov sveta do 108 *sémantických domén*, ktoré vypracoval E. A. Nida (1975). Naše delenie vzniklo (a vzniká) na princípe kontrastívnej diferenciácie; je teda ospravedlnené tým, že funkčnosť parafrázovača si zatiaľ nevyžiadala potrebu jemnejšieho delenia.

Problém zaraďovania do sémantických tried

Pri klasifikácii lexém do sémantických tried sme museli riešiť problém, na ktorý narazil aj E. A. Nida (ibid., s. 189). Štrukturácia sémantických domén je diskutabilná a hranice medzi nimi sú neostré.

Ak ide o polysémne lexémy jedného slova, môžeme ich klasifikovať každú samostatne:

diamant$_1$	[+mat]	(drahokam)
diamant$_2$	[+ins,+artef]	(nástroj na rezanie skla)

Ak nie je vhodné pripustiť polysémiu, lexéma dostane zoznam alternatívnych príznakov. Napríklad *piesok* môže vystupovať vo funkcii materiálu (na stavbe), ale aj aktívneho prírodného živlu (na púšti) a inštrumentu (brúsny piesok na brúsenie zrkadiel):

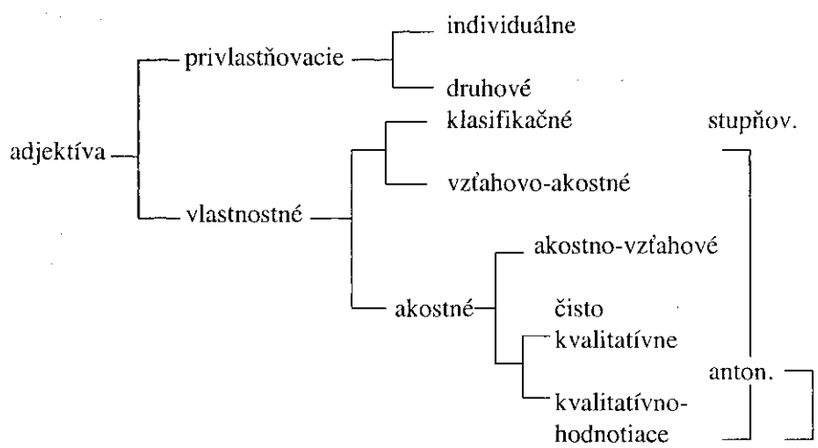
piesok$_1$	[+mat;+sila;+ins]

Na ten istý problém upozorňuje J. Pustejovsky (1991), a uvádza príklad substantív typu *kniha, videokazeta*, ktoré sa môžu chápať raz ako informačný obsah média, inokedy ako fyzické médium samotné. Túto ambivalenciu taktiež rieši zoznamom alternatívnych príznakov.

Dištinktívne príznaky adjektív

Triedenie adjektív preberáme z knihy Súčasný slovenský jazyk, Morfológia (Oravec - Bajzíková - Furdík, 1988). Nad odvodenými adjektívami vzniká ešte sekundárne triedenie podľa črty podstatného mena (resp. slovesa), od ktorého sú

odvodené. Tieto dve triedenia spolu sú vzhľadom na princíp kontrastívnej diferenciácie dostatočné.



Dištinktívne príznaky slovies

Dištinktívnym sémantickým príznakom slovesa je číslo jeho intenčného rámca, resp. triedy (5.10, 5.11). Napríklad:

[1] translačné intranzitívne vedomé procesy
[2] translačné intranzitívne nevedomé procesy
[20] konštruktívny manuálny proces
[50] percepčný proces orientovaný
[60] verbálny proces
[80] kognitívny proces s akuzatívnym objektom
[400] verba essendi

Príslušným číslom označený intenčný rámec potom obsahuje množinu sémantických pádov, aké daný typ slovies vytvára:

[1] [pro,des$_1$,des$_7$,ins,via$_1$,mds,des$_8$,mot,ori$_7$,kaz,knd,des$_{12}$]
[60] [edi,frm,ads,tem,sub,ins,ben,res,ads]
[80] [kog,fen,mds,ins,ben]

Na základe princípu kontrastívnej diferenciácie (0.16) vzniká bohaté triedenie slovies, ktoré vyplýva z ich obsažnej lexikálnej sémantiky a syntakticko-väzobnej mnohotvárnosti. V (Páleš, 1988) sme uviedli 120 slovesných intenčných tried, ktoré vznikli klasifikáciou slovies z Krátkeho slovníka slovenského jazyka (1987).

Dištinktívne príznaky prísloviek a druhotných predložiek

Sémantickým príznakom prísloviek a druhotných predložiek je priamo sémantický pád (5.14, 5.16), ktorý *vždy* vytvárajú. Význam prísloviek ako aj druhotnopredložkových fráz je viac-menej nezávislý od kontextu a teda ku každej príslovke, resp. druhotnej predložke možno uviesť konštantný sémantický pád:

dôstojne	[mds]
presne	[mds]
veľa	[mes]
dole	[lok_3]
nadol	[des_3]
zdola	[ori_3]
dolom	[via_3]
denne	[fre]
odvčera	[inc]
dozajtra	[fin]
spoza	[ori_5]
vôkol	[lok_6]
naprieč	[via_1]
kvôli	[mot]
za účelom	[mot]
vďaka	[kaz]
s ohľadom na	[res_1]
na úkor	[ben_2]

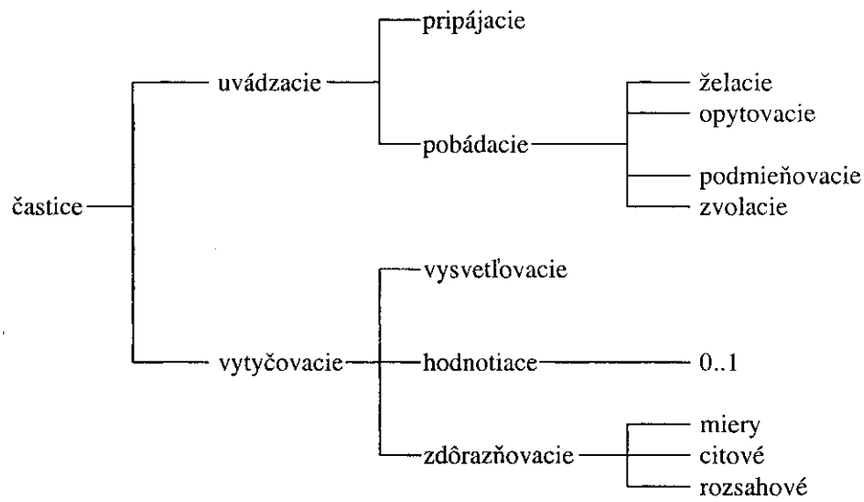
Dištinktívne príznaky spojok, častíc a citosloviec

Spojky, častice a citoslovcia sú v slovníku charakterizované dvojicou sémantických príznakov. Parataktické spojky sú rozlíšené druhom vzťahu, do akého uvádzajú vety v súvetí. Špecifické hypotaktické spojky sú charakterizované sémantickým pádom, do ktorého stavajú podradenú vetu. Význam všeobecných hypotaktických spojok (*že, čo, ktorý...*) sa vyhraňuje až v kontexte a rieši sa až počas syntakticko-sémantickej analýzy. Častice a citoslovcia sú taktiež rozdelené podľa svojich druhov a funkcií. Pri hodnotiacich časticiach je uvedené číslo z intervalu <0,1> podľa toho, akú veľkú mieru istoty tieto častice vyjadrujú.

parataktické spojky	zlučovacie	súčasné deje	(simult.)
		následné deje	(sekven.)
		striedavé deje	(altern.)
	odporovacie	vlastné	(adverzat.)
		želacie nesplnené	(deziderat.)
		obmedzovacie	(restring.)
		výnimka	(except.)
		konfrontačné	(konfron.)
		kompenzačné	(kompenz.)
		prípustkové	(konces.)
	vylučovacie	ostro vylučovacie	(disjunkt.)
		eventualitné	(event.)
		vylučovacia otázka	(quest.)
		opravujúce	(rektif.)
	stupňovacie	gradačné	(grad.)
		degradačné	(degrad.)

hypotaktické spojky	všeobecné		(atr)
	špecifické	časové	(tmp)
		príčinné	(kaz)
		účelové	(mot)
		podmienkové	(knd)
		prípustkové	(kcs)
		spôsobové porovn.	(rel)
		účinkové	(kzk)
		zreteľové	(res)
		prostriedkové	(ins)

citoslovcia	citové	kladné
		záporné
	vôľové	upozorňovacie
		imperatívne
	zvukomalebné	pohybové
		zvukové



1.4 Slovník syntaktických väzieb

obsahuje informáciu o väzobných a kolokačných prostriedkoch jazyka, ktoré sú viac-menej formálnou záležitosťou alebo záležitosťou úzu. Táto informácia je pre syntaktický modul nepostrádateľná tak pri analýze, ako aj pri syntéze. Ide o:

Gramatické pády predložiek. Každá predložka sa viaže s jedným určitým gramatickým pádom, niektoré s dvoma alebo troma:

za	[gen, ak, ins]
o	[ak, lok]
nad	[ak, ins]
z	[gen]
k	[dat]
s	[ins]

Pádové väzby slovies. Každé sloveso (slovesná trieda) sa viaže s určitou špecifickou množinou holých gramatických alebo predložkových pádov. Napríklad sloveso *vyrobiť* patrí do intenčnej triedy 20 a prislúcha mu zoznam:

[0+nom,0+ak,0+dat,0+ins,z+gen,na+lok,pre+ak,na+ak,za+gen...]

Zoznam gramatických pádov v syntakticko-väzobnom slovníku zodpovedá zoznamu sémantických pádov v lexikálno-sémantickom slovníku a spolu tvoria intenčný rámec slovesa:

[0+nom, 0+ak, 0+dat, o+lok, cez+ak, ...] *(povedať)*
[edi, frm, ads, tem, ins, ...]

Pádové väzby substantív. Deverbatívne substantíva spravidla preberajú intenciu slovies, ktorými sú motivované. Hoci nie vždy je to tak. Napríklad vo dvojici *milovať-láska*, *láska* priberá predložkový pád *k+datív*, ktorý sa pri pôvodnom slovese *milovať* nenachádza. Je to dané tým, že po nominalizácii sa nemôžu agens aj paciens *milovania* súčasne zobraziť do genitívu. Aj nemotivované substantíva môžu priberať špecifické väzby, ktoré treba uviesť:

láska	[k+dat]
povesť	[o+lok]
chuť	[+inf]

Dvojice parataktických spojok. Niektoré parataktické spojky utvárajú stabilné dvojice so špecifickým významom tak, že jedna zo spojok uvádza prvú vetu v súvetí a druhá druhú. Syntaktický analyzátor musí správny spoluvýskyt takýchto dvojíc kontrolovať:

len-hneď, sotva-už	... následné deje
to-to, tu-tu, raz-a_zase	... striedavé deje
síce-zato	... kompenzačné deje
buď-buď, buď-alebo	... disjunktné deje
či-či	... vylučovacia otázka

1.5 Slovník slovotvorných hniezd

zoraďuje lexiku do slovotvorných hniezd. Obsahuje teda informáciu o vzťahoch odvodenosti medzi jednotlivými slovami. Vedomie týchto vzťahov je pre identifikáciu slovotvornej zložky lexikálneho významu dôležité.

Pri slovotvornej analýze a syntéze je dôležitá nielen informácia o tom, že nejaká dvojica slov je vo vzťahu odvodenosti, ale často je cenná aj informácia o tom, že jedno slovo od druhého odvodené *nie je*. Napríklad, že slová *streľák, streľič, streľník* nie sú deriváty od slovesa *strieľať*, *ryba* nie je dejové deverbatívum od *ryť*; *škorica, mačka, čajka* nie sú prechýlené od *škorec, mak, čajník* a pod. Príklady slovotvorných hniezd *sviet-, kráľ-* a *zlat-* sú uvedené na nasledujúcej strane.

Pri zoraďovaní slovnej zásoby do slovotvorných hniezd sme narazili na rad ťažkostí. Ide o problémy, na ktoré upozorňujú K. Buzássyová (1974), J. Horecký (1971) a Mluvnice češtiny (1986). Pri programovaní systému SAPFO sme sa s každým z nich museli nejakým heuristickým spôsobom vysporiadať.

Problém organizácie slovotvorného hniezda. V niektorých prípadoch nie je zrejmé, od ktorého z viacerých možných koreňov je dané slovo odvodené. Napr. *hojdačka* by mohlo byť odvodené od zvratného *hojdať sa* rovnako, ako od *hojdať*. V iných prípadoch nie je jasný smer slovotvornej motivácie. Napr. pri vidových dvojiciach slovies alebo dvojiciach ako *neha-nežný* alebo *blázon-blázniť*.

Takéto prípady sme posudzovali a riešili z hľadiska pragmatického, t.j. z hľadiska schopnosti počítača tvoriť parafrázy. Zvolili sme takú štrukturáciu slovotvorného hniezda, aby bola čo najviac v súlade so všeobecnými pravidlami formálnej gramatiky systému a umožňovala tvorbu žiadúcich parafráz. V prípade, keď je voľba z hľadiska tvorby parafráz irelevantná, sme si jednoducho zvolili pravidlo, podľa ktorého sme uprednostňovali jednu z alternatív, a toto pravidlo sme sa snažili uplatňovať systematicky. Tak sme napríklad v role slovotvorného základu uprednostňovali nedokonavé slovesá pred dokonavými, nezvratné pred zvratnými, frekventované pred menej frekventovanými, gramaticky a štylisticky bezpríznakové pred príznakovými.

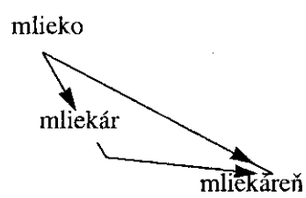
Slovotvorné hniezda *sviet-, kráľ-, zlat-*.

svietiť	kráľ
svietidlo	kráľovstvo
svieca	kráľovná
sviečka	kráľovič
sviečkár	kráľovský
sviečkáreň	kraľovať
sviečkový	veľkráľ
sviečkovitý	
svietnik	
svetlo	zlato
svetielko	zlatý$_1$
svetielkovať	zlatistý
svetlica	zlatisto
svetelný	zlato
svetelnosť	zlatka
svetlomet	zlatný
svetluška	zlatučký
Svetlana	zlatulinký
Svetlonoš	zlatokrídly
svetlocitlivý	zlatovlasý
osvetliť	zlatovláska
osvetľovať	zlatník
presvetliť	zlatníčka
presvetľovať	zlatníctvo
svietivý	zlatnícky
svietivosť	zlatotepec

svetlý	zlatotepectvo
svetlastý	zlatotepecký
svetlučký	zlatokop
svetlooký	zlatobyľ
svetlovláska	zlatonosný
svetlobelasý	zlatoočko
zosvetliť	zlatožltý
osvietiť	zlatočervený
osvetľovač	zlátiť
osvecovať	pozlátiť
poosvecovať	pozlacovať
osvietenie	popozlacovať
osvietenec	pozlátka
presvietiť	zlatieť
presvecovať	zlatoň
zasvietiť	
rozsvietiť	
rozsvecovať	
rozsvietiť sa	
dosvietiť	

Problém viacnásobnej motivácie. K jednému slovu môžu existovať súčasne dve alebo viac derivačných ciest, z ktorých všetky môžu byť prijateľné. Napríklad české slovo *soudce* môže byť odvodené od substantíva *soud* a zároveň od slovesa *soudit*. *Mladica* môže byť derivované priamo od prídavného mena *mladý*, ale aj prechýlené od *mladík*.

Za dominantnú môžeme zvoliť tú slovotvornú motiváciu, ktorá sa pociťuje ako silnejšia alebo ktorá zodpovedá frekventovanejšiemu slovotvornému postupu. Druhá možnosť je ponechať v databanke obe derivačné cesty súčasne:



V takom prípade si potom počítač 'uvedomuje' obe slovotvorné motivácie a konštruuje aj oba alternatívne druhy parafráz:

<u>mliekáreň</u>
miesto, kde je mlieko
miesto, kde sú mliekári

Problém polysémickej diferenciácie významu. Dynamické pohyby v slovnej zásobe spôsobujú, že po čase už význam lexém nemusí zodpovedať významu, aký by naznačovala slovotvorná forma a v nej použité formálne prostriedky. Ide najmä o významotvorné postupy lexikalizácie a polysémickej diferenciácie.

Polysémickú diferenciáciu riešime tak, že ju necháme odrážať sa v diferenciácii (osamostatňovaní) slovotvorných hniezd. Napríklad slovo *srdce* - z hľadiska čisto formálnej derivatológie - tvorí 21-členné hniezdo:

srdce
- srdiečko
  - srdiečkový
- srdcový
- srdečný
  - srdečne
  - srdečnosť
- srdnatý
- srdcovitý
  - srdcovito
  - srdcovite
- milosrdný
  - milosrdenstvo
  - milosrdnosť
- srdcelomný
- srdcervúci
- srdciar
- srdcovnica
  - srdcovnicový
- osrdcovník
- srdcovka

Takáto formálna štrukturácia slovotvorného hniezda je však úplne slepá k významu. Disparátne významy sú zmiešané dohromady, čo by nevyhnutne vniesolo zmätok do parafrázovania. Formálne členenie napríklad ignoruje, že adjektíva *srdečný*, *srdcový* a *srdcovitý* sú odvodené každé od iného významu slova *srdce*.

Pretože SAPFO nepracuje s lexémami, ale s lexiami, t. j. s každým významom slova ako s osobitným slovom, každá lexia si prirodzeným spôsobom utvára svoje vlastné slovotvorné hniezdo, ktoré je s ňou aj významovo konzistentné. V slovníku sa nachádza päť významov slova *srdce:*

$srdce_1$	korp	(telesný orgán)
$srdce_2$	abstr	(schopnosť cítiť)
$srdce_3$	vec	(predmet srdcovitého tvaru)
$srdce_4$	lok	(centrálne položené miesto)
$srdce_5$	ins	(srdce zvonu)

K nim prislúchajú samostatné hniezda:

srdce$_1$	srdce$_2$	srdce$_3$	srdce$_4$
srdiečko$_1$	srdečný	srdcovitý	
srdcový$_1$	srdečne	srdcovito	
srdciar	srdečnosť	srdcovite	
srdcovnica	milosrdný	srdiečko$_3$	
srdcovnicový	milosrdenstvo	srdiečkový	
osrdcovník	milosrdnosť	srdcovka	
	srdcervúci		
	srdcelomný		srdce$_5$
	srdiečko$_2$		
	srdcový$_2$		

Štruktúra slovotvorného slovníka SAPFO teda nie je formálne-slovotvorná, ale slovotvorne-významová. Ak sú dve slová umiestnené v jednom slovotvornom hniezde, znamená to, že zachovávajú slovotvorný význam. Ak ho nezachovávajú, t. j. došlo k nejakému významovému posunu, diferenciácii či lexikalizácii, posunutá lexia sa vyčleňuje s celým svojím podhniezdom ako samostatné hniezdo, alebo je podhniezdo označené značkou, ktorá na druh významového posunu upozorňuje.

Vďaka tomu slovotvorný slovník pomáha pri odstraňovaní nejednoznačností počas syntaktickej analýzy. Prívlastok *srdečný* automaticky poukazuje na abstraktum *srdce*$_2$ ako ľudský cit, pretože v inom hniezde sa nenachádza. Podobne *srdcovitý* rieši polysémickú štvorznačnosť jednoznačne v prospech konkréta *srdce*$_3$, poukazujúc na tvarovú podobnosť.

Problém lexikalizácie a idiomatizácie slovotvorného významu. Ďalším problémom je posun významu po lexikalizácii. Väčšina slov svoj slovotvorný význam zachováva. Napr. *strelec* je ten, kto strieľa; *slepec* je ten, kto je slepý. No nie vždy je to tak. Napr. *staviteľ* je nie len osoba, ktorá stavia domy, ale aj osoba, pre ktorú je stavanie povolaním (porovnaj *stavár, stavbár).* Substantíva *vodák, vodník, vodár* sú derivované od *voda* synonymnými sufixami, no každé má svoj osobitý význam.

Pokiaľ lexikálny význam slova jeho slovotvorný význam nepopiera, ale iba zužuje, resp. rozširuje, slová ponecháme v spoločnom slovotvornom hniezde. To je v súlade s horeuvedeným pravidlom, že v jednom hniezde majú byť tie a len tie slová, ktoré zachovávajú slovotvorný význam. Slovotvorný modul je napokon zodpovedný *iba* za slovotvornú časť významu slova. Ostatné zložky lexikálneho významu sú záležitosťou ostatných modulov, ktoré k analýze významu slova prispievajú a spoločne ho skladajú. Tak napríklad *zlatka* nie je všetko, čo je zlaté, ale iba zlatá minca. Na druhej strane *stolár* nevyrába iba stoly, ale nábytok všeobecne. Na cizelácii významu sa v prípade *zlatky* zúčastní slovník hyperoným a v prípade *stolára* konotačný a lexikálno-paradigmatický slovník (porov. príklady 7.10, 7.15).

V prípade, keď je slovotvorný význam mierne posunutý, je na mieste posunu v slovotvornom hniezde značka, ktorá slovotvorný analyzátor upozorňuje, že treba

hľadať alternatívne, spravidla slabšie interpretácie. Napríklad *zelenina, modrina* nemusia byť zelené (resp. modré) vždy, ale iba v typickom prípade. Adjektívum *zlatý* sa nachádza v jednom slovotvornom hniezde dvakrát v dvoch významoch, z ktorých jeden je označený ako posunutý:

zelený	
zelina	(vždy zelená)
zimozel	(vždy zelený)
~zelenina	(spravidla zelená)
zlato	
zlatý$_1$	(vyrobený zo zlata)
~zlatý$_2$	(má len farbu ako zlato)
zlatý$_3$	(vzácny)

V krajnom prípade dôjde k úplnej demotivácii a strate slovotvornej zložky významu. Také slovo musí tvoriť samostatné hniezdo (napr. *jed-jedlo).*

Supletívne a cudzojazyčné korene. Ako sme videli, štruktúra slovotvorného slovníka SAPFO nie je formálno-slovotvorná, ale významovo-slovotvorná. Preto do slovotvorných hniezd neváhame zaradiť ani slová derivované cudzími afixami, ani slová derivované od neslovenských koreňov. Rovnako ako pri supletívnom tvaroslovnom stupňovaní (pekný - krajší) alebo supletívnom skloňovaní (Zeus - Dia) pracujeme s nimi iba ako so supletívnymi koreňmi. Napríklad do nášho hniezda pre *srdce*$_1$ pribudne ďalších šesť slov, pretože grécky koreň *kard-* je významovo ekvivalentný slovenskému koreňu *srdc-* v prvom význame:

srdce$_1$
srdiečko$_1$
srdcový$_1$
srdciar
kardiak
srdcovnica
  srdcovnicový
osrdcovník
kardiológia
  kardiológ
  kardiologický
kardiostimulátor
srdcovocievny
kardiovaskulárny

V angličtine, ktorá má zdanlivo iba rudimentárnu slovotvorbu, je práve tento spôsob tvorenia slov dominantný. Vzťahové adjektíva a iné deriváty sa odvodzujú

anglickými afixami od prevzatých koreňov, najčastejšie latinských, talianskych, francúzskych a i.:

to carry	→	portable	(portare = to carry, tal.)
moon	→	lunar	(luna = moon, lat.)
star	→	constellation	(la stella = star, tal.)
king	→	royal	(le roy = king, fr.)
shine	→	luminous	(lumen = light, lat.)
god	→	divine	(diva, deva = god, indoeur.)

Neexistujúce články slovotvorného reťazca. Nie všetky potenciálne možnosti slovotvorného systému sa aktuálne uplatnia v slovnej zásobe. Dôvod môže byť čisto historický, alebo taký, že štruktúra fyzikálneho univerza jednoducho nepripúšťa existenciu denotátu, ktorý by slovotvorná forma mohla označovať. Existuje sloveso *zhorknúť*, ale nie *zoslanúť*, pretože samo od seba sa nič nestáva slaným. Naproti tomu existuje *osladiť*, ale nie *zhorčiť*, lebo nikto si úmyselne nerobí jedlo horkejším. To zatiaľ nepredstavuje problém. SAPFO neexistujúce slová neanalyzuje ani netvorí, pretože slovotvorný syntezátor konzultuje svoje výsledky so slovotvorným slovníkom, ktorý poskytuje informáciu o tom, aké slová v slovenskej slovnej zásobe *skutočne* existujú a aké nie.

Problém vzniká, ak derivatologický reťazec existujúcich slov logicky predpokladá odvodzovací článok, ktorý však v slovnej zásobe chýba.

Napríklad *znárodniť* je odvodené od *národ*, ale medzičlánok *národniť* neexistuje. V tomto prípade medzičlánok ani nepokladáme za potrebný, lebo priamu (preskokovú) deriváciu *národ - znárodniť* sme na základe jej frekvencie povýšili na samostatný slovotvorný postup (analogický k *ticho - utíšiť sa*).

V ostatných prípadoch je rekonštrukcia derivačného medzičlánku nevyhnutná. Slovotvorné hniezda SAPFO takéto medzičleny skutočne obsahujú a pracujú s nimi.

Rozlišujeme pri tom dva prípady, podľa toho, či chýbajúci logický člen má, alebo nemá význam. V prvom prípade by mohol existovať, ale neexistuje z nejakých iných príčin. V druhom prípade neexistuje a ani by existovať nemohol, lebo ani denotát takého javu nemôže existovať. Napríklad v prípade *žena - zženštilý* uvádzame do hniezda medzičlen - sloveso *zženštiť sa*. Toto sloveso sa potom objaví aj pri parafrázovaní.

Druhý prípad predstavuje trojica *obuť - zobuť - vyzuť*, očividne tvoriaca slovotvornú rodinu, no vôbec nie je jasné, aký člen by mal byť ich spoločným otcom. Ide o tzv. *viazaný základ* (porov. Horecký, 1959). Spoločný koreň z pochopiteľných dôvodov neexistuje. Človek sa totiž vždy môže buď len *obuť*, alebo *zobuť*, no nikdy nie *iba uť*. Napriek tomu - podľa princípu jedno-jednoznačnosti reprezentácie (0.15) - musí existovať komponent významu, ktorý zodpovedá prieniku (abstrakcii) tohto slovotvorného hniezda. Nazveme ho *uť*. V slovotvornom hniezde je označený krížom - na znamenie, že tento komponent významu sa síce môže zúčastňovať sémantických inferencií, ale medzi povrchovými formami sa nikdy nesmie objaviť samostatne:

#uť
zobuť
vyzuť
obuť
obuv
obuvník
obuvák

Mŕtve slovotvorné postupy. Hoci niektoré slovotvorné postupy už nie sú produktívne, SAPFO s nimi pracuje ako so živými. Ide napríklad o zaniknuté l-príčastia a m-príčastia. Kým si slovo zachováva regulárny slovotvorný význam, patrí do slovotvorného hniezda bez ohľadu na to, že daný slovotvorný postup z diachronického hľadiska už nefunguje:

kameň
kameňovať
skamenieť
skamenelý
skamenelina

Keby sme napr. slovo *dospelý* chápali len ako neodvodené adjektívum, popierali by sme tým slovotvornú zložku jeho významu, a znemožnili parafrázy typu '*dospelý je ten, kto dospel*', '*nevidomý je ten, kto nevidí*'. To, že lexéma *dospelý* má ešte aj iné zložky významu, so slovotvornou analýzou nijako nesúvisí, ale je to už záležitosť lexikálno-sémantického modulu.

Keď príčastie slovotvornú zložku svojho významu stratilo, do slovotvorného hniezda samozrejme nepatrí (*skvieť sa - skvelý, smieť - smelý, z dvora - zdvorilý*).

1.6 Synonymický a lexikálno-paradigmatický slovník

Lexikálno-paradigmatický slovník je databankou paradigmatických vzťahov v lexike. Tieto sú zachytené vo forme tzv. *lexikálnych parametrov*. Lexikálny parameter je systematicky sa opakujúci vzťah v lexike. Niektoré naše parametre sú príbuzné, alebo priamo prevzaté od J. D. Apresiana (1974) a I. A. Meľčuka (1974).

Do paradigmatických vzťahov vstupujú lexie, nie lexémy. Preto napr. *vzácny*$_1$ je vo vzťahu synonymie iba so *zlatý*$_3$, nie so *zlatý*$_1$ alebo *zlatý*$_2$.

Slovník lexikálnych parametrov tvorí informačný základ pre celý rad syntaktico-sémantických transformácií (12.6).

Do paradigmatického slovníka sú zaradené aj vzťahy významovo paralelné so slovotvorbou, ktoré však vzhľadom na svoju formu do slovotvorby nepatria. Napríklad *učiteľ* je AGENSom *učenia* a tento vzťah je indikovaný aj formálne, vďaka čomu ho slovotvorný analyzátor má možnosť identifikovať. *Žiak* je

PACIENSom *učenia*, ale vzťah medzi touto dvojicou slov nie je indikovaný formálne. Kedysi existovalo slovo *učeník*, naznačujúce obsah aj svojou formou. Vzťahy ako *zlodej - kradnúť*, *vrah - zabiť* musia byť vymenované ako súčasť lexikálno-paradigmatického slovníka.

Príklad. Časť lexikálno-paradigmatického slovníka

dublet(trocha$_1$,trochu$_1$).

synon$_1$(krídlo$_1$,peruť$_1$).
synon$_1$(loďstvo$_1$,flotila$_1$).
synon$_1$(mesiac$_1$,luna$_1$).
synon$_1$(čaša$_1$,kalich$_1$).

synon$_2$(zlatý$_3$,vzácny$_1$).
synon$_2$(čarokrásny$_1$,divukrásny$_1$).

synon$_3$(kúzliť$_1$,čarovať$_1$).

anton$_1$(deň$_1$,noc$_1$).

anton$_2$(drahý$_1$,lacný$_1$).
anton$_2$(pekný$_1$,škaredý$_1$).
anton$_2$(plný$_1$,prázdny$_1$).

konverz(kúpiť$_1$,predať$_1$).
konverz(darovať$_1$,obdarovať$_1$).
konverz(ukradnúť$_1$,okradnúť$_1$).

inverz(naliať$_1$,vyliať$_2$).
inverz(priniesť$_1$,odniesť$_1$).

synekdocha(stôl$_1$,nábytok$_1$).

usual$_1$(korunovať$_1$,slávnostne$_1$).
usual$_2$(ovácia$_1$,búrlivá$_1$)

verbal$_1$(čarovať,robiť čary).
verbal$_1$(báť sa,mať strach).

nominaliz(cestovať$_1$,na+lok).

dublet(okno$_1$,oblok$_1$).

magnon$_1$(pršať$_1$,liať$_5$).

magnon$_1$(biely$_1$,snehobiely$_1$).
magnon$_1$(červený$_1$,ohnivočervený$_1$).
magnon$_1$(krásny$_1$,čarokrásny$_1$).
magnon$_1$(krásny$_1$,divukrásny$_1$).

magnon$_2$(strach$_1$,panický$_1$).
magnon$_2$(ticho$_1$,mŕtvy$_1$).
magnon$_2$(pohnevať$_1$,na_smrť$_1$).
magnon$_2$(zhorieť$_1$,na_popol$_1$).
magnon$_2$(zhorieť,do_tla)

symbol$_1$(láska$_1$,srdce$_2$).
symbol$_2$(ľstivý$_1$,líška$_1$).

kompar(pomalý$_1$,slimák$_1$).
kompar(starý$_1$,Matuzalem$_1$).

bonif(kúpiť$_1$,výhodne$_1$).
bonif(strieľať$_1$,presne$_1$).

abstr$_1$(orol$_1$,vták$_1$).
abstr$_1$(náhrdelník$_1$,šperk$_1$).
abstr$_2$(ebenový$_2$,drevený$_1$).
abstr$_3$(odliať$_3$,vyrobiť$_1$).
abstr$_4$(vprostriedku,v)

verbal$_1$(pršať,padať dážď).
verbal$_2$(šepkať,hovoriť ticho).

nominaliz(jesť$_1$,pri+lok).

reduk$_1$(osram$_1$,osmium$_1$-volfrám$_1$).
reduk$_1$(motel$_1$,motoristický$_1$ hotel$_1$).
reduk$_1$(moped$_1$,motorový$_1$ velociped$_1$).
reduk$_2$(matika$_1$,matematika$_1$).
reduk$_2$(anglina$_1$,angličtina$_1$).
reduk$_3$(socdem$_1$,sociálna$_1$ demokracia$_1$)
reduk$_3$(kominterna$_1$,komunistická$_1$ internacionála$_1$)
reduk$_4$(kilo$_1$,kilogram$_1$).
reduk$_4$(auto$_1$,automobil$_1$).
reduk$_4$(Magda$_1$,Magdaléna$_1$).
reduk$_5$(Betka$_1$,Alžbetka$_1$).
reduk$_6$(telka$_1$,televízia$_1$).
reduk$_6$(profák$_1$,profesor$_1$).

abrev$_1$(tzv.,takzvaný$_1$).
abrev$_1$(str.,strana$_2$).
abrev$_1$(prof.,profesor$_1$).
abrev$_2$(čs.,československý$_1$).
abrev$_2$(kpt.,kapitán$_1$).
abrev$_2$(rtg.,röntgenový$_1$).
fon(GAZ$_1$,gazík$_1$).
fon(ZNB$_1$,esenbé$_1$,).
fon(PKO$_1$,pekáč$_1$).
apo(kebys',keby si).
apo(čos',čo si).
apo(hoc',hoci).

akron(FF$_1$,filozofická$_1$ fakulta$_1$).
akron(USA$_1$,Spojené$_1$ štáty$_1$ americké$_1$).
akron(ZRPŠ$_1$,združenie$_1$ rodičov$_1$ a priateľov$_1$ školy$_1$).
akron(UK$_1$,Univerzita$_1$ Komenského$_1$).
akron(UK$_2$,Univerzita$_1$ Karlova).

val(1,jeden$_1$).
val(3,tri$_1$).
val(12,dvanásť$_1$).
val(12,tucet$_1$).
val(100,sto).
sign(+,plus$_1$).
sign(%,percento$_1$).
sign(&,a$_1$).
sign(V,volt$_1$).
sign(cos,kosínus$_1$).

hyperb(malý,malililinký).
hyperb(dlhý,dlhááánsky).
združ(vysoká škola).
združ(vlčí mak).

Viacslovné výrazy sa v slovníku hneď transformujú na hĺbkové štruktúry. Významové manipulácie, unifikácia a inferencia sa potom robia so sémantickými sieťami, nie s povrchovými formami:

funkc([zlodej / «AGS kradnúť]).

funkc([žiak / «PAC učiť / «AGS učiť sa]).

verbal(čarovať, [robiť / REZ» čary]).

verbal(pršať, [padať / PRO» dážď]).

verbal(báť sa, [mať / OBJ» strach]).

verbal(šepkať, [hovoriť / ATR» ticho]).

$$\text{akron(USA,}\begin{bmatrix}\text{štát}_1\text{-} \\ \text{«PAC spoj}_1\text{-} \\ \text{LOK» Amerik}_1\text{-} \\ \text{+pl}\end{bmatrix}\text{).}$$

$$\text{akron(ZRPŠ,}\begin{bmatrix}\text{združ}_1\text{-} \\ \text{AGS» rodič}_1\text{-} \\ \text{+pl} \\ \text{AGS» osob}_1\text{-} \\ \text{«AGS priatel}_1\text{-} \\ \text{+pl} \\ \text{«AGS škol}_2\text{-}\end{bmatrix}\text{)}$$

$$\text{akron(FF,}\begin{bmatrix}\text{fakult}_1\text{-} \\ \text{OKU» filozof}_1\text{-}\end{bmatrix}\text{).}$$

1.7 Frazeologický slovník

Frazeologický slovník obsahuje informáciu, ktorú využíva frazeologický modul programu na analýzu, syntézu a parafrázovanie frazeologických zvratov. O spôsobe ich reprezentácie a použitia pozri kapitolu 9.

Príklad. Časť frazeologického slovníka

frazeol(od_výmyslu_sveta,všelijak-).
frazeol(kosť_a_koža,chud-).

$$\text{frazeol(}\begin{bmatrix}\text{meč-} \\ \text{LDET» Damokles-}\end{bmatrix}\text{,}\begin{bmatrix}\text{nebezpečenstv-} \\ \text{«AGS hroz-} \\ \text{MDS» ustavičn-}\end{bmatrix}\text{)}$$

$$\text{frazeol(}\begin{bmatrix}\text{klinec-} \\ \text{LDET» zlat-}\end{bmatrix}\text{,}\begin{bmatrix}\text{vyvrcholen-}\end{bmatrix}\text{).}$$

$$\text{frazeol(}\begin{bmatrix}\text{nalia-} \\ \text{BEN» X} \\ \text{0+GEN» vín-} \\ \text{ATR» čist-}\end{bmatrix}\text{,}\begin{bmatrix}\text{hovor-} \\ \text{OBJ» pravd-}\end{bmatrix}\text{)}$$

$$\text{frazeol(}\begin{bmatrix}\text{mať-} \\ \text{+neg} \\ \text{AGS» X} \\ \text{OBJ» srdc-}\end{bmatrix}\text{,}\begin{bmatrix}\text{byť} \\ \text{ANT» X} \\ \text{ATR» bezcitn-} \\ \text{+expr}\end{bmatrix}\text{)}$$

$$\text{frazeol(}\begin{bmatrix}\text{padn-} \\ \text{BEN» X} \\ \text{DO+GEN» ok-}\end{bmatrix}\text{,}\begin{bmatrix}\text{zapáč- sa} \\ \text{AFE» X}\end{bmatrix}\text{)}$$

$$\text{frazeol(}\begin{bmatrix}\text{vyhráv-} \\ \text{AGS» cigán-} \\ \text{+pl} \\ \text{BEN» X} \\ \text{LOK}_1\text{» bruch-}\end{bmatrix}\text{,}\begin{bmatrix}\text{mať} \\ \text{AGS» X} \\ \text{OBJ» hlad-} \\ \text{+mag}\end{bmatrix}\text{)}$$

$$\text{frazeol}\left(\begin{bmatrix} X & \\ & \text{ZA+GEN» groš-} \\ & \text{DET» judášsk-} \end{bmatrix}, \begin{bmatrix} X & \\ & \text{KAZ» ziskuchtivosť-} \\ & \text{+expr} \end{bmatrix}\right).$$

1.8 Štylistický slovník

Štylistický slovník obsahuje informáciu o štylistickej, terminologickej a nárečovej príznakovosti lexiky, ako aj o jej pôvode. Túto informáciu využíva štylistický modul programu.

Aj štylistické príznaky sú priradené lexiám, nie lexémam. Napríklad *vyletieť*$_2$ je slangové, *vyletieť*$_1$ spisovné.

Štylistickú hodnotu slova pritom pokladáme za konštantnú, t. j. nepočítame s tým, že slovo môže rozličné štylistické hodnoty nadobudnúť umiestnením v kontexte.

Nepracujeme s pojmom neologizmu, pretože je veľmi diskutabilný.

Príklad. Časť štylistického slovníka

štyl(sťa$_1$,poet).
štyl(deva$_1$,poet).
štyl(luna$_1$,poet).

štyl(kiež$_1$,arch).
štyl(slnce$_1$,arch).
štyl(krasopanna$_1$,arch).
štyl(meruôsmy$_1$,arch).

štyl(blahoslavený$_1$,bibl).

štyl(dievka$_1$,kniž).
štyl(zakúpiť$_1$,kniž).
štyl(zásluhou$_1$,kniž).
štyl(následkom$_1$,kniž).

štyl(ekvipment$_1$,fam).
štyl(Pišta$_1$,fam).
štyl(zicherhajcka$_1$,fam).

štyl(oháknuť sa$_1$,slang).
štyl(esenbák$_1$,slang).
štyl(vyletieť$_2$,slang).
štyl(prachy$_1$,žarg).

proven(filozofia$_1$,gréc).

štyl(deci$_1$,hovor).
štyl(počuj$_1$,hovor).
štyl(človeče$_1$,hovor).

štyl(kosínus$_1$,odb).
štyl(difluóretán$_1$,odb).
štyl(fluorit$_1$,odb).

štyl(dolupodpísaný$_1$,admin).
štyl(t.r.$_1$,admin).
štyl(dostaviť sa$_1$,admin).

štyl(zosnúť$_1$,eufem).
štyl(žrať$_1$,pejor).
štyl(tetuška$_1$,laud).
štyl(nosisko$_1$,augm).

štyl(počuj$_1$,explet).
štyl(všakáno$_1$,explet).
štyl(človeče$_1$,explet).

dial(sľiuka,turiec).

proven(cisár$_1$,lat).
proven(kostým$_1$,franc).
proven(skúter$_1$,angl).
proven(šachta$_1$,nem).
proven(guláš$_1$,maď).
proven(kulak$_1$,rus).
proven(chémia$_1$,arab).
proven(azúr$_1$,perz).

dial(šecia,liptov).
dial(kämeň,novohrad).
dial(lokec,šariš).
dial(dziuka,zemplín).
dial(slifka,trnava).
dial(kóň,zvolen).
dial(kúň,záhorie).
dial(dzesec,spiš).

termin(determinant,mat).
termin(determinant,lingv).
termin(izokyanid,chem).
termin(heliograf,astron).

termin(fermentácia,biochem).
termin(feromagnetický,fyz).
termin(systolický,lek).
termin(arpeggio,hud).

1.9 Slovník asociatívnych väzieb a konotácií

Asociatívny slovník zachytáva nepovinné, ale typicky sa opakujúce kolokácie (Osgood a kol., 1957).

Napríklad pri niektorých podstatných menách existujú špecifické slovesné deje, v ktorých dané podstatné meno spravidla vystupuje ako aktant v niektorej sémantickej role (Pustejovsky, 1991). K niektorým substantívam takéto doplnenia existovať nemusia, k niektorým ich môže existovať viac, ale žiadne nemusí byť špecifické.

O asociatívnych väzbách možno uvažovať medzi ľubovoľnými slovnými druhmi a SAPFO ich využíva hlavne na rekonštrukciu informácie stratenej po nominalizačných, derivačných a kompozičných postupoch.

Asociatívny slovník môže vzniknúť manuálne alebo frekvenčnou analýzou korpusu textov (Smadja, 1993).

Príklad. Časť ascociatívneho slovníka

asoc([víno «PAC piť]).
asoc([víno «REZ dorobiť]).
asoc([chlieb «PAC jesť]).
asoc([chlieb «REZ piecť]).
asoc([kvitnúť PRO» kvet]).

asoc([ihla «INS šiť]).
asoc([nôž «INS rezať]).
asoc([ok-$^{+pl}$ ATR» modr-]).
asoc([ok-$^{+pl}$ ATR» hned-]).
asoc([vojak «AGS bojovať]).

2.

FONOLÓGIA

Ak chceme formálny model jazyka vybudovať od základov, musíme začať popisom hláskoslovia. SAPFO obsahuje úplnú klasifikáciu foném podľa ich vlastností prevzatú od J. Dvončovej (1980) a J. Horeckého (1977).

Triedenie hlások podľa ich zvukových a artikulačných vlastností, ako aj informácia o paradigmatických vzťahoch medzi nimi (hlásky párové podľa dĺžky, tvrdosti a znelosti) sú potrebné a využívajú sa pri tvaroslovnej a slovotvornej analýze a syntéze. Pri časovaní, skloňovaní a odvodzovaní vznikajú koreňové alternácie, ktorých prípustnosť a typ závisí práve od fonologického okolia slovotvorného alebo tvarotvorného švíka.

Napríklad pri vkladaní vokálu v genitíve plurálu ženských vzorov (*matka* → *matiek, vojna* → *vojen, jamka* → *jamôk, perla* → *perál, kráska* → *krások*) musí morfológia konzultovať modul fonológie. Alternácia je prípustná, ak sa koreň končí konsonantickou skupinou dvoch tvrdých spoluhlások. Podľa toho, či je predchádzajúca slabika krátka alebo dlhá, bude vložený dlhý alebo krátky vokál (rytmický zákon). Od ďalších vlastností zúčastnených konsonantov (sykavosť, sonórnosť) závisí, či vložený vokál bude mať *e*-ové, *a*-ové alebo *o*-ové zafarbenie.

Klasifikácia foném

FONÉMY (GRAFÉMY)
podľa prítomnosti tónovej zložky
VOKÁLY (samohlásky)
podľa dĺžky
KRÁTKE: a, e, i, y, o, u, v
DLHÉ: á, é, í, ý, ó, ú, ia, ie, iu, ô
podľa rezonančných priestorov
ZADNÉ (velárne): o, ó, u, ú
STREDNÉ: a, á
PREDNÉ (palatálne): e, é, i, í, y, ý
VYSOKÉ: i, í, y, ý, u, ú
STREDOVÉ: e, é, o, ó
NÍZKE: a, á, ä
podľa zaokrúhlenia pier
LABIALIZOVANÉ (zaokrúhlené): u, y
NELABIALIZOVANÉ (nezaokrúhlené): i, e, v

podľa otvorenosti pier
OTVORENÉ: a, á
POLOOTVORENÉ: e, é, o, ó, v
ZATVORENÉ: i, í, y, ý, u, ú
DIFTONGY (dvojhlásky): ia, ie, iu, ô
POLOVOKÁLY: v
SONÓRY: r, l, ŕ, Ĺ
KONSONANTY (spoluhlásky)
podľa dĺžky
KRÁTKE: r, l
DLHÉ: ŕ, Ĺ
podľa tvrdosti
TVRDÉ: h, ch, k, g, d, t, n, l
MÄKKÉ: c, č, dz, dž, f, j, š, ž, ť, ď, ň, ľ
OBOJAKÉ: v, m, r, b, p, s, v, z
podľa účasti hlasu
ZNELÉ: b, d, ď, g, dz, dž, z, ž
NEZNELÉ: p, t, ť, k, c, č, s, š
PÁROVÉ: b+p, d+t, ď+ť, g+k, dz+c, dž+č, z+s, ž+š
NEPÁROVÉ: r, l, m, n, ň, j
podľa miesta artikulácie
PERNOPERNÉ (bilabiálne): p, b, m
PERNOZUBNÉ (labiodentálne): f, v
PREDOĎASNOVÉ (prealveolárne): t, d, n, s, z, c, dz
ZADOĎASNOVÉ (postalveolárne): š, ž, č, dž, r, ŕ, l, Ĺ
ĎASNOVOPODNEBNÉ (alveopalatálne): ť, ď, ň, ľ
TVRDOPODNEBNÉ (palatálne): j
MÄKKOPODNEBNÉ (velárne): k, g, ch
HRTANOVÉ (laryngálne): h
podľa sluchového dojmu
EXPLOZÍVNE (výbuchové): p, b, m, t, d, n, ť, ď, ň, k, g
AFRIKÁTY (polotrené): c, č, dz, dž
FRIKATÍVY (trené): s, z, š, ž, h, ch, f, v, r, ŕ, l, Ĺ, ľ
SONÓRY: m, n, ň, j, r, l, ľ
LIKVIDY (plynné): r, l, ŕ, Ĺ
VIBRANTY: r, ŕ

3.

MORFOLÓGIA

3.1 Potreba morfologickej analýzy

V analytických jazykoch s rudimentárnou morfológiou sa v praxi uplatnia aj *systémy bez morfologickej analýzy* (Wilensky a kol., 1988). Napríklad morfológiu angličtiny možno riešiť slovníkovo, t. j. vytvorením zoznamu všetkých existujúcich tvarov spolu s ich gramatickými kategóriami.

Vo flektívnych jazykoch takýto postup nie je únosný. V prípade slovenčiny by sa objem slovníka zväčšil cca 40x (Páleš, 1989).

3.2 Definícia morfologickej analýzy a syntézy

Uskutočniť *morfologickú analýzu* znamená určiť k danému slovnému tvaru všetky jeho gramatické kategórie. Opačným procesom je morfologická *syntéza*. Znamená utvorenie slovného tvaru na základe daných gramatických kategórií.

Analýzu a syntézu tvarov vykonávajú predikáty *substant, adjekt, adverb, indikat, imperat, lpar, ntpar, cpar, tranz, gerund, infin, prepoz, conj, partik, interj*. Každý z nich je špecializovaný na jeden okruh tvarov a má dve skupiny argumentov - gramatické a sémantické:

```
substant(gram(Rod,Čís,Pád,Zápor),sem(Sém,Životn,Počít,Vzor),Ko,X).
adjekt(gram(Rod,Čís,Pád,Živ,Stup,Záp),sem(Druh,SklVz,StpVz),Ko,X).
adverb(gram(Stupeň,Záp),sem(Druh),Ko,X).

indikat(gram(Osob,Čís,Čas,Záp),sem(Intencia,Zvrat,Vid),Km,X).
imperat(gram(Osob,Čís,Záp),sem(Intencia,Zvrat,Vid),Km,X).
lpar(gram(Rod,Čís,Záp),sem(Intencia,Zvrat,Vid),Km,X).
ntpar(gram(Rod,Čís,Pád,Živ,Stup,Záp),sem(Intencia,Zvrat,Vid),Km,X).
cpar(gram(Rod,Čís,Pád,Živ,Čas,Záp),sem(Intencia,Zvrat,Vid),Km,X).
tranz(gram(Záp),sem(Intencia,Zvrat,Vid),Km,X).
gerund(gram(Pád,Čís,Záp),sem(Intencia,Zvrat,Vid),Km,X).
infin(gram(Záp),sem(Intencia,Zvrat,Vid),Km,X).

prepoz(gram(Pád,Vokalizovanosť),sem(Druh),Ko,X).
conj(gram(),sem(Taktickosť,Druh1,Druh2,Počet,Cst),Ko,X).
partik(gram(),sem(Druh1,Druh2),Ko,X).
interj(gram(),sem(Druh1,Druh2),Ko,X).
```

Predikáty sú reverzibilné, t. j. možno ich použiť obojsmerne: pre analýzu aj pre syntézu. Ak za posledný argument dosadíme nejaký slovný tvar, program doplní jeho gramatické kategórie. A naopak, ak zadáme gramatické kategórie, program vygeneruje žiadaný tvar:

```
?-lpar(gram(Rod,Čís,Záp),sem(Intencia,Zvrat,Vid),Kmeň,čítala).
Rod=žen; Čís=sg; Záp=poz;
Intencia=60; Zvrat=nezvr;
Vid=nedok; Kmeň=čít;
yes.

?-ntpar(gram(muž,sg,dat,než,sup,poz),_,čít,X).
X=najčítanejšiemu.
yes.
```

Výsledok analýzy býva spravidla viacznačný. Za viacznačnosť zodpovedá tvarová homonymia spolu s funkčnou homonymiou tvarových prípon. Úlohou analyzátora je nájsť *všetky* alternatívne analýzy. Ich ďalšie zjednoznačnenie je úlohou syntaxe:

```
?-analyzuj(X,moria).
X=substant(gram(str,pl,nom,poz),_,more,moria);
X=substant(gram(str,pl,akuz,poz),_,more,moria);
X=indikat(gram(3os,pl,prít,poz),_,moriť,moria);
yes.

?-analyzuj(X,nežnú).
X=adjekt(gram(žen,sg,akuz,živ,poz,poz),_,nežný,nežnú);
X=indikat(gram(3os,pl,prít,neg),_,žať,nežnú);
yes.
```

3.3 Procedúra analýzy

Slovný koreň/kmeň so systémom pádových/časovacích prípon spolu s typom alternácie tvoria postačujúcu informáciu pre generovanie ľubovoľného tvaru. Nie však pre analýzu.

Teoreticky, ak vieme tvary generovať, mali by sme ich vedieť aj analyzovať. To je pravda. To však nestačí na to, aby bol proces analýzy zvládnuteľný aj z hľadiska výpočtovej zložitosti. Systém potrebuje ďalšiu, dodatočnú a veľmi podrobnú informáciu, ak sa analýza má stať cieľavedomým, efektívnym procesom, a nie slepým prehľadávaním.

Povedzme, že chceme zanalyzovať nejaký tvar. Napríklad *„najnepochopenejším"*. Doplníme ho ako posledný argument a ostatné argumenty necháme neviazané, aby ich systém doplnil:

?-analyzuj(X,najnepochopenejším).

PROLOG na túto výzvu odpovedá exhaustívnym generovaním všetkých možných tvarov od všetkých slovenských koreňov a porovnáva ich s analyzovaným tvarom.

Problém spočíva v tom, že ak máme pred sebou neznámy tvar, nevieme ešte, aký tvar máme analyzovať; ba nevieme ani o aký slovný druh pôjde (lebo to má byť až výsledkom analýzy) - a preto nevieme ani akým spôsobom, „z akého konca" začať analýzu.

Exhaustívna metóda zrejme nie je reálna. (Počet slovenských tvarov sa blíži k desiatkam miliónov).

3.4 Izolácia koreňa

Pretože v slovníku nie sú uvedené celé slová, ale len korene a slovesné kmene, systém, aby mohol v slovníku identifikovať koreň, musí ho najprv izolovať od tvarových predpôn a prípon.

V slovenskom slove možno oddeliť 0-5 znakov ako predponu, 0-4 znaky ako skloňovaciu (resp. časovaciu) príponu a 0-3 znaky ako relačnú príponu (pri slovesách). To dáva spolu 5.4.3 = 60 možností rozštvrtenia tvaru na koreň a afixy. Keďže o koreni nevieme, či nie je pozmenený niektorou z 30-tich druhov alternácií, musíme sa pokúsiť odalternovať ho. Tým dostávame 60.30 = 1800 hypotetických koreňov pri jednom jedinom tvare ! Z týchto 1800 možností bude perspektívnych iba niekoľko, prieskum ostatných je zbytočná nadpráca.

Musíme teda postupovať inak. Vybavíme SAPFO dôkladnou znalosťou fonematických a morfematických spájateľnostných zákonitostí slovenčiny. To mu umožní sformulovať vždy iba niekoľko pravdepodobných hypotéz, možných prípadov, o aký tvar ide a od akého koreňa by asi mohlo ísť. Ak takýto koreň naozaj v slovníku nájde, overí si svoj predpoklad ešte generatívne, t. j. vyskloňovaním alebo vyčasovaním. V prípade, že dostane pôvodne analyzovaný tvar, hypotéza bola správna.

3.5 Hypotézy o rozdelení na koreň a afixy

Slovenské tvarotvorné afixy sa nekombinujú ľubovoľne, ale ich spájateľnosť je veľmi špecifická. Navyše z veľkého počtu hláskových kombinácií nie všetky, ale iba niektoré tvoria slovenský afix.

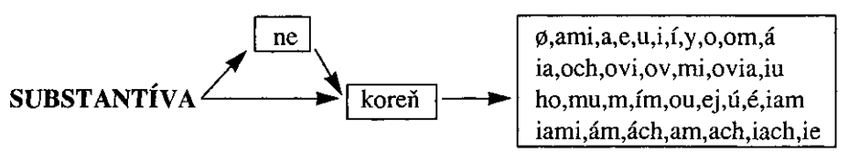
SAPFO využíva skutočnosť, že slovenské tvary nemôžu byť zakončené ľubovoľným koncovým podreťazcom, ale iba podreťazcami z istej konečnej množiny. Ak slovo má byť napr. podstatným menom, musí byť zakončené niektorou z 36-tich substantívnych prípon. Ak to má byť navyše podstatné meno v určitom rode, čísle

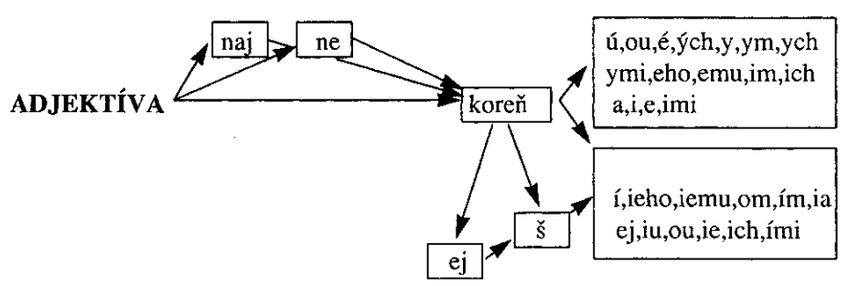
či páde, musí byť zakončené príponou, ktorá je pre tieto gramatické kategórie charakteristická (pozri tabuľky na nasledujúcich stranách).

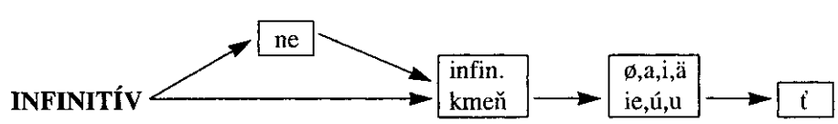
Ak by analyzovaný tvar mal byť napríklad l-ovým slovesným tvarom, musí byť zakončený na *l,la,lo* alebo *li*. Inak ním nie je. Nie naopak; tvar sa môže končiť na *l,la,lo,li* a l-ovým slovesným tvarom byť nemusí. Preto je hypotézu potrebné overiť.

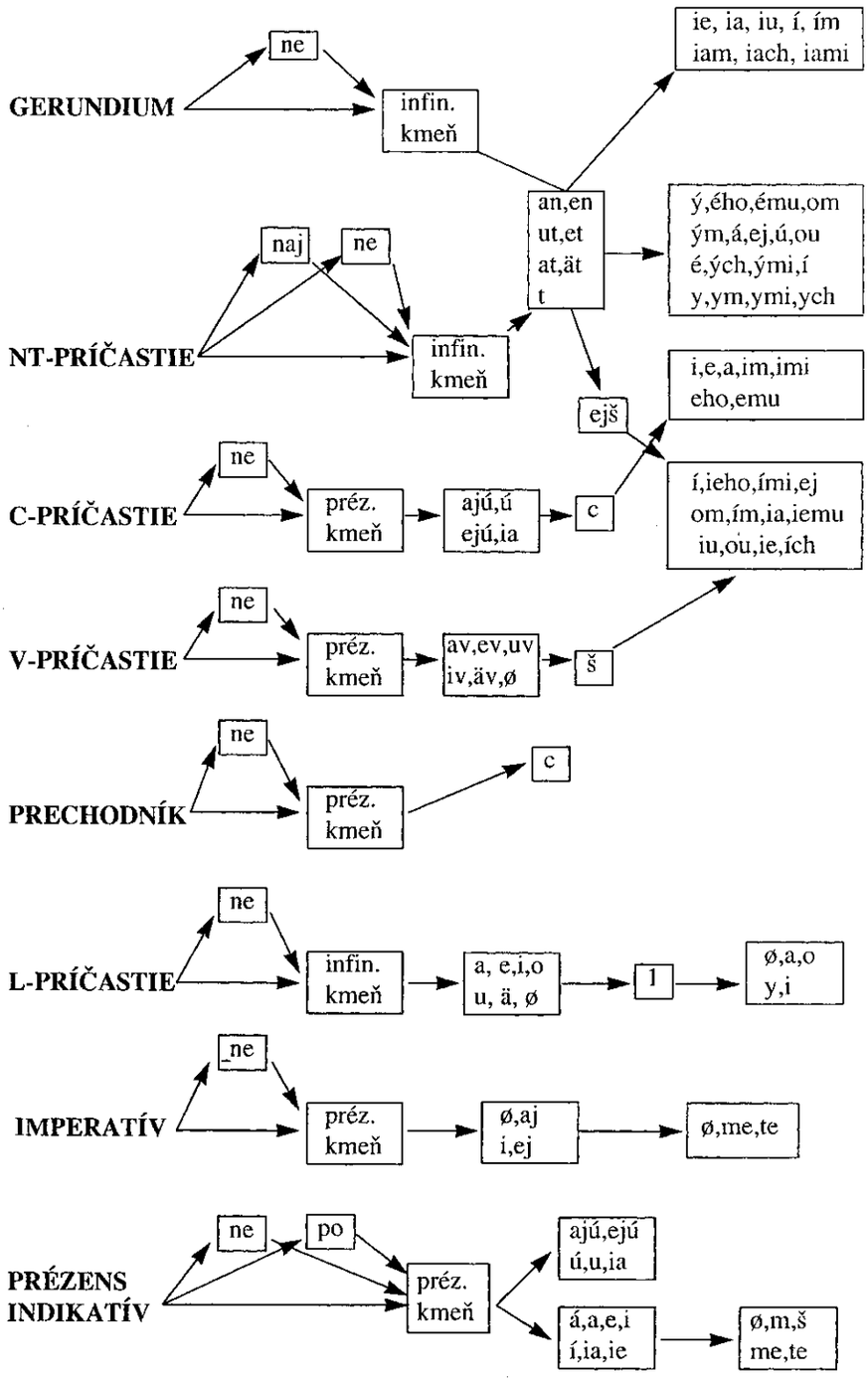
Teda morfologický modul SAPFO sa správa ako konečný nedeterministický automat, ktorého vstupným reťazcom sú fonémy analyzovaného tvaru (sprava doľava smerom od konca slova). Podľa vstupujúcich foném sa automat dostáva do stavov, ktoré mu ukladajú overiť len malý počet hypotéz-možností (teda už nie všetky). Obdobným spôsobom sa postupuje zľava doprava od začiatku slova pri oddeľovaní prefixov. Stavy automatu možno znázorniť grafmi a tabuľkami:

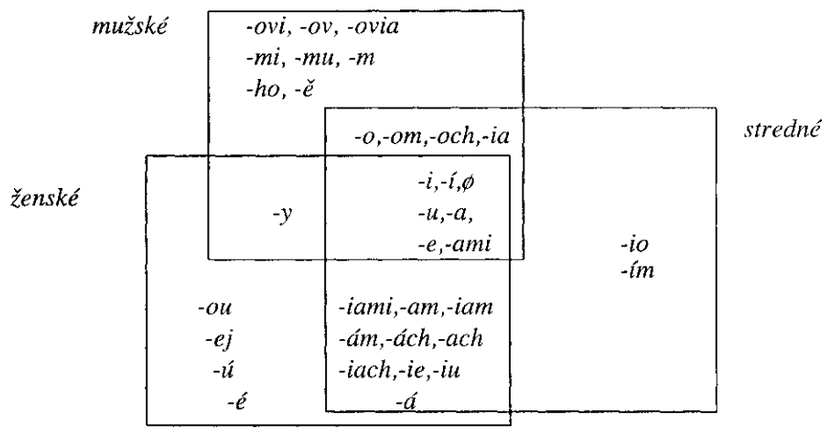
Spájateľnosť koreňov a tvarotvorných afixov



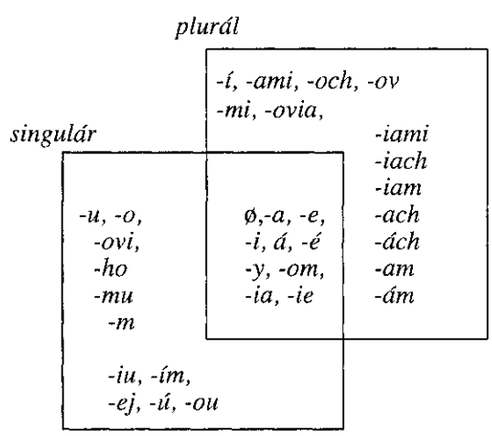





GERUNDIUM
ne
infin. kmeň
ie, ia, iu, í, ím iam, iach, iami
NT-PRÍČASTIE
naj
ne
infin. kmeň
an,en ut,et at,ät t
ý,ého,ému,om ým,á,ej,ú,ou é,ých,ými,í y,ym,ymi,ych
ejš
i,e,a,im,imi eho,emu
C-PRÍČASTIE
ne
préz. kmeň
ajú,ú ejú,ia
c
í,ieho,ími,ej om,ím,ia,iemu iu,ou,ie,ích
V-PRÍČASTIE
ne
préz. kmeň
av,ev,uv iv,äv,ø
š
PRECHODNÍK
ne
préz. kmeň
c
L-PRÍČASTIE
ne
infin. kmeň
a, e,i,o u, ä, ø
l
ø,a,o y,i
IMPERATÍV
ne
préz. kmeň
ø,aj i,ej
ø,me,te
PRÉZENS INDIKATÍV
ne
po
préz. kmeň
ajú,ejú ú,u,ia
á,a,e,i í,ia,ie
ø,m,š me,te



Pádové prípony substantív a ich rodová príznačnosť



Pádové prípony substantív a ich príznačnosť podľa čísla

pád	*príznačné*	*polyfunkčné sufixy*
nominatív	ovia	ø,a,e,i,í,y,o,ia,é,á,ie
genitív		ø,a,e,u,i,í,y,ia,ov,ho,ej
datív	mu,ám,am,iam	ø,e,u,i,om,ovi,ej,iu
akuzatív	ú	ø,a,e,u,i,í,y,o,ia,ov,ho,á,iu,ie
lokál	och,ách,ach,iach	ø,e,u,i,í,om,ovi,m,ej
inštrumentál	ou,ím,mi,ami,iami	ø,om,m

Pádové prípony substantív a ich príznačnosť podľa pádov

Minimálny automat na separáciu koreňov (sprava doľava)								
0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*sufix príznačný pre tvar*
ø	.	.	.	.	.	.	.	sub, adj, prind, imper, neoh.
	a	.	.	.	.	.	.	sub, adj, ntpar
		c	a	i	.	.	.	cpar
			á	.	.	.	.	cpar
			ú	j	a	.	.	cpar
					e	.	.	cpar
		i	.	.	.	.	.	sub, adj, prind, ger
			š	.	.	.	.	vpar
				v	.	.	.	vpar
			v	o	.	.	.	sub
		l	.	.	.	.	.	lpar
	á	.	.	.	.	.	.	sub, adj, ntpar
	c	i	a	.	.	.	.	tranz
		ú	.	.	.	.	.	tranz
			j	a	.	.	.	tranz
				e	.	.	.	tranz
	e	.	.	.	.	.	.	sub, adj
		c	a	i	.	.	.	cpar
			á	.	.	.	.	cpar
			ú	j	a	.	.	cpar
					e	.	.	cpar
		i	.	.	.	.	.	sub, adj, ger
			š	.	.	.	.	vpar
				v	.	.	.	vpar
		m	.	.	.	.	.	prind, imper
			i	.	.	.	.	imper
			j	a	.	.	.	imper
				e	.	.	.	imper
		t	.	.	.	.	.	prind, imper
			i	.	.	.	.	imper
			j	a	.	.	.	imper
				e	.	.	.	imper
	é	.	.	.	.	.	.	sub, adj, ntpar
	h	c	a	.	.	.	.	sub
				i	.	.	.	sub, ger
			á	.	.	.	.	sub
			i	.	.	.	.	adj
				c	a	i	.	cpar
					á	.	.	cpar
					ú	j	a	cpar
							e	cpar
			í	.	.	.	.	adj
				š	.	.	.	vpar
					v	.	.	vpar
			o	.	.	.	.	sub

0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*sufix príznačný pre tvar*
			y	.	.	.	.	adj
			ý	.	.	.	.	adj, ntpar
	i	.	.	.	.	.	.	sub, adj, imper, ntpar
		c	a	i	.	.	.	cpar
			á	.	.	.	.	cpar
			ú	j	a	.	.	cpar
					e	.	.	cpar
		l	.	.	.	.	.	lpar
		m	.	.	.	.	.	sub
			a	.	.	.	.	sub
				i	.	.	.	sub, ger
			i	.	.	.	.	adj
				c	a	i	.	cpar
					á	.	.	cpar
					ú	j	a	cpar
							e	cpar
			í	.	.	.	.	adj
				š	.	.	.	vpar
					v	.	.	vpar
			y	.	.	.	.	adj, ntpar
			ý	.	.	.	.	adj, ntpar
		v	o	.	.	.	.	sub
	í	.	.	.	.	.	.	sub, adj, ntpar, ger
		š	.	.	.	.	.	vpar
			v	.	.	.	.	vpar
	j	a	.	.	.	.	.	imper
		e	.	.	.	.	.	sub, adj, imper, ntpar
			c	a	i	.	.	cpar
				á	.	.	.	cpar
				ú	j	a	.	cpar
						e	.	cpar
			š	.	.	.	.	vpar
				v	.	.	.	vpar
	l	.	.	.	.	.	.	lpar
	m	.	.	.	.	.	.	sub, prind
		a	.	.	.	.	.	sub
			i	.	.	.	.	sub, ger
		á	.	.	.	.	.	sub
		i	.	.	.	.	.	adj, ntpar
			c	a	i	.	.	cpar
				á	.	.	.	cpar
				ú	j	a	.	cpar
						e	.	cpar
		í	.	.	.	.	.	sub, adj, ger
			š	.	.	.	.	vpar
		o	.	.	.	.	.	sub, adj, ntpar
			c	a	i	.	.	cpar
				á	.	.	.	cpar

0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*sufix príznačný pre tvar*
				ú	j	a	.	cpar
						e	.	cpar
			š	.	.	.	.	vpar
				v	.	.	.	vpar
		y	.	.	.	.	.	adj, ntpar
		ý	.	.	.	.	.	adj, ntpar
	o	.	.	.	.	.	.	sub, adj, prep
		h	.	.	.	.	.	sub, adj
			e	.	.	.	.	adj, ntpar
				i	.	.	.	adj
					š	.	.	vpar
						v	.	vpar
			é	.	.	.	.	adj, ntpar
		l	.	.	.	.	.	lpar
	ô	.	.	.	.	.	.	adj
	š	.	.	.	.	.	.	prind
	ť	.	.	.	.	.	.	inf
		a	.	.	.	.	.	inf
		ä	.	.	.	.	.	inf
		e	i	.	.	.	.	inf
		i	.	.	.	.	.	inf
		u	.	.	.	.	.	inf
		ú	.	.	.	.	.	inf
	u	.	.	.	.	.	.	sub, adj, prind, ntpar, prep
		c	a	i	.	.	.	cpar
			á	.	.	.	.	cpar
			ú	j	a	.	.	cpar
					e	.	.	cpar
		i	.	.	.	.	.	sub, adj, ger
			š	.	.	.	.	vpar
				v	.	.	.	vpar
					a	.	.	vpar
					ä	.	.	vpar
					e	.	.	vpar
					i	.	.	vpar
					u	.	.	vpar
		m	.	.	.	.	.	sub
			e	.	.	.	.	adj, ntpar
				c	a	i	.	cpar
					á	.	.	cpar
					ú	j	a	cpar
							e	cpar
				i	.	.	.	adj
					š	.	.	vpar
						v	.	vpar
			é	.	.	.	.	adj, ntpar
		o	.	.	.	.	.	sub, adj, ntpar
			c	a	i	.	.	cpar

0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*sufix príznačný pre tvar*
				á	.	.	.	cpar
				ú	j	a	.	cpar
						e	.	cpar
			š	.	.	.	.	vpar
				v	.	.	.	vpar
	ú	.	.	.	.	.	.	sub, adj, prind, ntpar
	ú	j	a	.	.	.	.	prind
			e	.	.	.	.	prind
	v	o	.	.	.	.	.	sub
	y	.	.	.	.	.	.	sub, adj, ntpar
		l	.	.	.	.	.	lpar
	ý	.	.	.	.	.	.	adj, ntpar

Minimálny automat na separáciu koreňov (zľava doprava)						
0	*1*	*2*	*3*	*4*	*5*	*prefix príznačný pre tvar*
ø	n	a	j	.	.	adj, adv, ntpar
				n	e	adj, adv, ntpar
		e	.	.	.	sub, adj, slovesné tvary
			p	o	.	prind
	p	o	.	.	.	prind

3.6 Príklad

Ukážme si na príklade, čo sa deje pri analýze tvaru *vila*. Podľa stavov automatu (v tabuľke) môže ísť o:

1) substantívum s príponou ø a koreňom *vila*
2) substantívum s príponou *a* a koreňom *vil*

Substantívne koncovky *-la*, *-ila* už neexistujú. Ak krátke *i* v kmeni vzniklo skrátením samohlásky (alternáciou), ide o:

3) substantívum s príponou ø a koreňom *víla*
4) substantívum s príponou *a* a koreňom *víl*

ďalej môže ísť o:

5) adjektívum s príponou ø a koreňom *vila*

6) adjektívum s príponou *a* a koreňom *vil*

7) určité sloveso s tematickou morfémou *a* a časovacou príponou ø od slovesného kmeňa *vil*

8) l-ový slovesný tvar s tematickou morfémou *i* a časovacou príponou *-la* od slovesného kmeňa *vi*

Týmto sme vylúčili možnosť, že by *vila* mohol byť prechodník alebo niektoré z činných príčastí (neobsahuje *š* ani *c*). Vylúčený je neurčitok (zakončenie *ť*). Nemôže ísť ani o trpné príčastie alebo slovesné podstatné meno.

Teda ako hypotetické korene prichádzajú do úvahy:

vila, vil, víla, víl, vi

Slovník potvrdí iba existenciu koreňov:

vil (od podstatného mena *vila)*
vi (od slovesa *viť)*
víl (od ženského podstatného mena *víla)*

Z týchto troch hypotéz sa nakoniec spätne potvrdia iba prvé dve. Substantívum *víla*, ktoré sa skloňuje podľa vzoru *dáma*, nemá predpokladanú alternáciu (skrátenie samohlásky) uvedenú, teda *vila* nie je žiaden tvar od *víla*. Celková analýza vyzerá:

```
?- analyzuj(vila,X)
X = subst(vila,vil,nom,sg,žen,neživ,konkr,žena);
X = l-príčastie(vila,viť,žen,sg,pozit,nezvrat,nedok,piť);
yes
```

3.7 Hypotézy o alternáciách

Samotným oddeľovaním prípon hľadaný koreň nikdy nenájdeme v prípade, že je alternovaný. Napríklad pri oddeľovaní prípon od tvaru *turkami* dostávame

turkami + 0, turkam + i, turka + mi, turk + ami

nie však skutočný koreň, ktorý znie *turek*.

V tomto prípade SAPFO ďalej generuje hypotézy o eventuálnej alternácii a tvare pôvodného koreňa. Po preskúmaní fonologickej štruktúry prichádzajú do úvahy len štyri typy alternácií. Pretože *turk* je zakončené dvojicou spoluhlások, program sa medzi ne pokúsi vložiť *e* alebo *o*. Pritom narazí na skutočný koreň, ktorý znie *turek*:

Alternovaný koreň	Pôvodný koreň (hypotéza)	Typ alternácie	
turkam	turkám	skrátenie vokálu	(mráz → mrazu)
turkam	turka	rozšírenie o ***m***	(ako Noe → Noe**ma**)
turk	turok	vypustenie ***o***	(čl**o**k → čl**n**ka)
turk	turek	vypustenie ***e***	(ot**e**c → ot**c**a)

3.8 Minimálna nevyhnutná nedeterministickosť

Morfologický automat SAPFO je nedeterministický. Avšak miera jeho nedeterministickosti je minimálna možná. Ide tu o mieru zložitosti, ktorá je jazyku vlastná (inherentná), a nie je viac redukovateľná zefektívnením výpočtového postupu.

Inak povedané, všetky výpočty, ktoré náš morfologický automat vykonáva, sú nevyhnutné, a automat, ktorý vykonáva menšie množstvo výpočtov, nepokrýva celú škálu javov slovenského tvaroslovia.

Pri analýze tvaru *víl musíme* uvažovať alternáciu s *víl*, hoci na konci výpočtu vysvitne, že substantívum *víla* nepripúšťa skrátenie vokálu. Pri slove *nefrit nemôžeme nepreskúmať* možnosť, že by išlo o zápor slovesa, lebo vopred nevieme, či *frit* nie je slovesný kmeň. *Ľane* by mohlo vzniknúť z *ľaň* (ako v *lode - lod'*). Ale *ľaň* sa nenájde medzi substantívnymi koreňmi. *Hrom* pripomína inštrumentál podstatného mena so štruktúrou *hr+om*. Dokonca aj koreň *hr* existuje, ale nie substantívny.

Pri analýze takýchto tvarov sa *musí* vykonať určité minimálne množstvo výpočtov, ktoré je *nevyhnutné*, pretože vyplýva z *vnútornej zložitosti* (miery entropie) samotného jazykového systému.

3.9 Vysvetlivky

sub	substantívum
adj	adjektívum
adv	adverbium
prind	prézens indikatív
imper	imperatív
lpar	l-ový slovesný tvar
ntpar	trpné príčastie
cpar	činné príčastie prítomné
vpar	činné príčastie minulé
tranz	prechodník
ger	gerundium
inf	neurčitok
prep	predložka
neoh.	spojka, častica, citoslovce

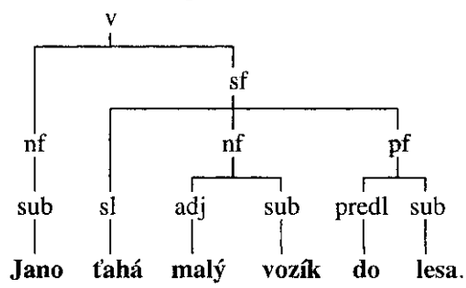
4.

SYNTAKTICKÁ ANALÝZA

4.1 Definícia syntaktickej analýzy

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

Výsledkom je *syntaktický (derivačný) strom*, ktorý vyjadruje vzťahy formálnej závislosti medzi frázami vo vete. Napríklad:



Veta *Jano ťahá malý vozík do lesa* sa skladá z mennej a slovesnej frázy. Menná fráza sa skladá z jednoduchého vlastného mena, slovesná fráza zo slovesa, ďalšej mennej a jednej predložkovej frázy. Menná fráza sa skladá z podstatného mena rozvitého prídavným menom, predložková z predložky a podstatného mena atď. Derivačné stromy sa generujú pomocou tzv. *prepisovacích pravidiel*. Pravidlá obsahujú neterminálne symboly, ktoré zastupujú všeobecné kategórie a frázy a sú označené malými písmenami. Terminálne symboly sú konkrétne slová (veľkými písmenami).

(1) v → nf,sf.
(2) sf → sl.
(3) sf → sl,nf.
(4) sf → sl,nf,pf.
(5) nf → sub.
(6) nf → adj,sub.
(7) nf → adj,sub,pf.
(8) pf → predl,sub.
(9) sub → JANO
(10) sub → VOZÍK
(11) sub → LESA
(12) adj → MALÝ
(13) sl → ŤAHÁ
(14) predl → DO

Neterminálne symboly:

sf	=	slovesná fráza
nf	=	menná, nominálna fráza
pf	=	predložková fráza
sub	=	substantívum
adj	=	adjektívum
sl	=	sloveso
v	=	veta

Terminálne symboly

Derivácia (odvodzovanie) sa začína *počiatočným symbolom* (v, veta). Ak sa v generovanom reťazci nachádza podreťazec neterminálnych symbolov, ktorý sa zhoduje s ľavou stranou nejakého pravidla, možno ho prepísať pravou stranou toho istého pravidla. Ak existuje viac pravidiel s rovnakou ľavou stranou, vzniká možnosť voľby. Odvodzovanie sa skončí nahradením celého reťazca samými terminálmi. To je odvodzovanie *zhora nadol*.

Pri odvodzovaní *zdola nahor* sa začína reťazcom samých terminálov a pravé strany pravidiel sa prepisujú ich ľavými stranami. Odvodenie sa skončí úspešne, ak sa reťazec podarí prepísať na počiatočný symbol (v).

Napríklad jedna z možných ciest odvodenia vety *Jano ťahá malý vozík do lesa* zhora nadol je:

		použité pravidlá
v →	nf,sf	(1)
→	nf,sl,nf,pf	(4)
→	sub,sl,nf,pf	(5)
→	sub,sl,adj,sub,pf	(6)
→	sub,sl,adj,sub,predl,sub.	(8)
→	JANO,sl,adj,sub,predl,sub	(9)
→	JANO,ŤAHÁ,adj,sub,predl,sub	(10)
→	...	(12),(10),(14)
→	JANO,ŤAHÁ,MALÝ,VOZÍK,DO,LESA	(11)

Možná cesta odvodenia vety *Jano ťahá malý vozík do lesa* zdola nahor:

	JANO,ŤAHÁ,MALÝ,VOZÍK,DO,LESA	
→	sub,ŤAHÁ,MALÝ,VOZÍK,DO,LESA	(9)
→	sub,sl,MALÝ,VOZÍK,DO,LESA	(13)
...		(12),(10),(14)
→	sub,sl,adj,sub,predl,sub	(11)
→	nf,sl,adj,sub,predl,sub	(5)
→	nf,sl,nf,predl,sub	(6)
→	nf,sl,nf,pf	(8)
→	nf,sf	(4)
→	v	(1)

4.2 Formálna gramatika

$$G = (V,N,T,P)$$

je teda daná štvoricou: počiatočným symbolom (V), množinou neterminálnych symbolov (N), terminálnych symbolov (T) a množinou pravidiel (P).

Na gramatiku sa možno dívať a skúmať ju z dvoch stránok:

(a) staticky: ako na formálnu špecifikáciu, vymedzenie prípustných gramatických štruktúr;

(b) dynamicky: ako na metódu, algoritmus analýzy a syntézy viet podľa danej gramatiky.

Pre implementáciu syntaktického analyzátora, ktorý má pracovať v reálnom čase, sú dôležité oba aspekty.

4.3 Chomského hierarchia

Každej gramatike sa dá priradiť jazyk (t. j. množina viet generovaných danou gramatikou) a naopak. Veľkosť a druh generovaného jazyka potom charakterizuje *generatívnu silu* a *senzitivitu* gramatiky.

Štyri najzákladnejšie typy gramatík (a im zodpovedajúcich jazykov) sú podľa generatívnej sily zoradené do tzv. *Chomského hierarchie*.

Regulárne jazyky sú generované regulárnou gramatikou, ktorá pozostáva výlučne z pravidiel typu $A \rightarrow aB$.

Bezkontextová gramatika (CFG, context free grammar) pozostáva z pravidiel s jedným neterminálom na ľavej strane (napr. $A \rightarrow aBc$).

Kontextové gramatiky (CSG, context-sensitive grammar) pripúšťajú pravidlá naľavo s viacerými neterminálmi za predpokladu, že počet symbolov na ľavej strane je menší alebo rovný počtu symbolov napravo ($AB \rightarrow aCd$, $AB \rightarrow BA$).

b Gramatika *typu 0* obsahuje pravidlá bez obmedzení, teda napr. aj prepis na prázdny symbol ($A \rightarrow \Sigma$).

Čo do generatívnej sily, sú gramatiky zoradené takto:

$$RG \leq CFG \leq CSG \leq \text{typu } 0$$

Podarilo sa ukázať (Hopcroft - Ullman, 1979), že existuje úzky vzťah medzi týmito štyrmi typmi jazykov a určitými druhmi automatických zariadení (formálnych automatov). Regulárne jazyky sa dajú generovať *konečnými automatmi*, bezkontextové *zásobníkovým automatom*, kontextové *konečne ohraničeným automatom* a jazyky typu 0 *Turingovými strojmi*:

formalizmus	jazyk
konečné automaty	regulárne jazyky
zásobníkové automaty	bezkontextové jazyky
ATN, DCG, HPSG, TAG, LFG	prirodzené jazyky
konečne ohraničené automaty	kontextové jazyky
Turingove stroje	jazyky typu 0

Hierarchia formálnych jazykov

Hierarchia formálnych jazykov a automatov bola vynikajúcim výsledkom syntaktického výskumu vo výpočtovej lingvistike 50-tych rokov. Čoskoro sa však ukázalo, že táto formálna klasifikácia nie je dostatočne vhodná pre opis prirodzených jazykov. Prirodzený jazyk totiž nie je bezkontextový, teda žiadna bezkontextová gramatika ho nemôže generovať v plnom rozsahu. Na druhej strane čistá kontextová gramatika prirodzené jazyky svojou generatívnou kapacitou mnohonásobne prevyšuje, takže vzniká opačný problém: ako vylúčiť vety, ktoré sú kontextovou gramatikou generované, ale do prirodzeného jazyka nepatria.

N. Chomsky (1965) sformuloval tézu, že kontextové gramatiky sú príliš silným a bezkontextové zase nedostatočným aparátom na zachytenie syntaxe prirodzeného jazyka. Úlohou výpočtovej lingvistiky bolo nájsť taký formálny aparát alebo automatické zariadenie, ktoré by svojou generatívnou silou zodpovedalo práve prirodzeným jazykom, t. j. negenerovalo by ani „príliš málo", ani „príliš veľa", ale práve množinu viet prirodzeného jazyka.

4.4 Bezkontextové gramatiky (CFG)

V počiatočnom štádiu výskumu sa pozornosť sústredila hlavne na bezkontextové gramatiky, ktoré možno zachytiť relatívne jednoduchým formalizmom a spracovávať zásobníkovými automatmi v polynomiálnom čase.

Bezkontextovú gramatiku možno použiť na syntézu a analýzu viet prirodzeného jazyka. Pri analýze sú možné postupy *zľava-doprava a sprava-doľava, zhora-nadol a zdola-nahor*. Z psycholingvistického hľadiska sa analýza vety sprava-doľava nepokladala za prirodzenú, pretože človek vníma jazyk vždy len zľava-doprava. Používala sa teda analýza zľava-doprava a to zhora-nadol, zdola-nahor alebo kombinovaným postupom.

4.5 Syntaktická analýza zhora-nadol

Začína sa počiatočným symbolom V (veta) a prepisuje ľavé strany pravidiel pravými, až kým nedostane samé terminálne symboly.

Dátové štruktúry:

- pozícia vo vstupnej vete (ktorá časť vety ešte nebola zanalyzovaná)
- reťazec odvodených symbolov, ktoré sa ešte neprekryli so vstupnou vetou
- návratové značky (markery) pri už použitých pravidlách alebo zásobník ešte nepoužitých pravidiel

Algoritmus:

Uvažujeme vždy najľavejší symbol v reťazci.

- Ak ide o neterminál, nájdeme pravidlo gramatiky s takýmto neterminálom na ľavej strane. Neterminál prepíšeme pravou stranou pravidla. Ak prichádza do úvahy viac pravidiel, ostatné možnosti uložíme do zásobníka.

- Ak ide o terminál a kryje sa so slovom na vstupe, odstránime symbol z reťazca a postúpime o 1 slovo vo vstupnej vete.

- Ak ide o terminál rôzny od terminálu na vstupe, znamená to, že zvolená odvodzovacia cesta je chybná a treba vykonať *proces návratu* (backtracking). Zo zásobníka sa vyberie jeden zo stavov, v ktorom sme mali na výber a pokračuje sa v analýze iným spôsobom. Ak je zásobník prázdny, celková analýza je neúspešná; veta nepatrí do jazyka generovaného danou gramatikou.

Analýza sa skončí úspešne, ak sa analyzátoru podarí dostať na koniec vstupného reťazca.

Ak pokračujeme uvedeným algoritmom až do vyprázdnenia zásobníka, môžeme dostať viacero úspešných alternatívnych spôsobov odvodenia jednej a tej istej vety.

Príklad:

$_1$ ***Jano*** $_2$ ***ťahá*** $_3$ ***vozík*** $_4$

krok	stav	zásobník stavov	pozícia	
1.	(v)		1	
2.	(nf sf)		1	
3.	(sub sf)		1	
		(adj sub sf)	1	
		(adj sub pf sf)	1	
4.	(sf)		2	kryje sa JANO
5.	(sl)		2	
		(sl nf)	2	
		(sl nf pf)	2	
6.	()		3	na vstupe zostal VOZÍK, neúspech
7.	(sl nf)		2	návrat k 5.
8.	(nf)		3	kryje sa ŤAHÁ
9.	(sub)		3	
		(adj sub)	3	
		(adj sub pf)	3	
10.	()		4	kryje sa VOZÍK

Koniec analýzy. Analýza sa skončila úspešne, pretože vstupný reťazec je prázdny () a hlavica automatu je v pozícii za posledným slovom vety (4).

V 6. kroku sa analyzátor pokúsil ukončiť analýzu s vetou JANO ŤAHÁ (podľa pravidla 2), ale na vstupe zostalo nezanalyzované slovo VOZÍK. Po návrate ku kroku 5 analýza pokračovala výberom tranzitívneho pravidla (3).

4.6 Syntaktická analýza zdola-nahor

Syntaktická analýza zdola-nahor sa začína terminálmi a snaží sa dostať počiatočný symbol. Prepisovacie pravidlá sa používajú opačným smerom, t. j. sprava doľava. Momentálne spracovávaný symbol sa nazýva *kľúč*. Analyzátor hľadá pravidlá gramatiky začínajúce sa kľúčom alebo už začaté pravidlá, ktoré potrebujú kľúč, aby sa dokončili alebo rozšírili. Pritom udržuje v pamäti tieto dynamické dátové štruktúry:

- zoznam konštituentov
- zoznam aktívnych pravidiel
- zoznam kľúčov, t. j. ukončených konštituentov, ktoré majú ešte vstúpiť do analýzy (ako súčasť vyšších konštituentov).

Keď je zoznam kľúčov prázdny, načíta sa ďalšie slovo zo vstupného reťazca a všetky jeho kategórie sa dajú do zoznamu kľúčov.

Algoritmus:

Ak vstúpi konštituent $C(p_1,p_2)$ v pozícii od p_1 po p_2:

- Pridaj $C(p_1,p_2)$ do zoznamu konštituentov.
- Ak s C začína pravidlo P v gramatike, pridaj P do zoznamu pravidiel ako aktívne od p_1 do p_2.
- Pre všetky aktívne pravidlá P, začínajúce sa v p_0 a končiace sa v p_1, ak C je ďalší konštituent v P, tak pridaj do zoznamu pravidiel nové aktívne pravidlo z p_0 do p_2, ktoré vznikne rozšírením P.
- Ak niektoré z aktívnych pravidiel v kroku 2 alebo 3 je ukončené, pridaj nový konštituent nazvaný ľavou stranou pravidla do zoznamu kľúčov.

Príklad:

Analyzujme nasledovnú anglickú vetu zdola-nahor pomocou gramatiky so šiestimi pravidlami:

$_1$ ***the*** $_2$ ***large*** $_3$ ***can*** $_4$ ***can*** $_5$ ***hold*** $_6$ ***the*** $_7$ ***water*** $_8$

(1)	S	←	NP VP	ART	article
(2)	NP	←	ART ADJ NOUN	ADJ	adjective
(3)	NP	←	ART NOUN	AUX	auxiliary verb
(4)	NP	←	ADJ NOUN	NP	noun phrase
(5)	VP	←	AUX VERB NP	VP	verb phrase
(6)	VP	←	VERB NP	S	sentence

Výsledok morfologickej analýzy je viacznačný a vyzerá takto:

the: ART
large: ADJ

can: AUX, NOUN, VERB
hold: NOUN, VERB
water: NOUN, VERB

Na začiatku je zoznam kľúčov prázdny.

()

Na zoznam kľúčov vstúpi prvé slovo a načne dve pravidlá:

→ *the (1,2)*

(ART)	začne	NP	←	ART+ADJ NOUN (1,2)
		NP	←	ART+NOUN (1,2)

()

→ *large (2,3)*

(ADJ)	začne	NP	←	ADJ+NOUN (2,3)
	posunie	NP	←	ART ADJ+NOUN (1,3)

Aktívne pravidlá sa nikdy z listiny neodstraňujú (neskôr by sa mohli použiť iným spôsobom).

→ *can (3,4)*

(AUX NOUN VERB)	začne	VP	←	AUX+VERB NP (3,4)
(NOUN VERB)	dokončí	NP	←	ADJ NOUN+ (2,4)
		NP	←	ART ADJ NOUN+ (1,4)
	tie otvoria	S	←	NP+VP (1,4)
		S	←	NP+VP (2,4)
(VERB)	otvorí	VP	←	VERB+NP (3,4)

→ *can (4,5)*

(AUX NOUN VERB)	otvorí	VP	←	AUX+VERB NP (4,5)
(NOUN VERB)		0		
(VERB)	otvorí	VP	←	VERB+NP (4,5)
	rozšíri	VP	←	AUX VERB+NP (3,5)

→ *hold (5,6)*

(NOUN VERB)		0		
(VERB)	otvorí	VP	←	VERB+NP (5,6)
	rozšíri	VP	←	AUX VERB+NP (4,6)

→ *the (6,7)*

(ART)	otvorí	NP	←	ART+ADJ NOUN (6,7)
		NP	←	ART+NOUN (6,7)

→ *water (7,8)*
(NOUN, VERB) ukončí NP ← ART NOUN+ (6,8) → medzi kľúče
(NP(6,8), VERB) ukončí VP ← VERB NP+ (5,8) → medzi kľúče
VP ← AUX VERB NP+ (4,8)
(VP(5,8), VP(4,8), VERB) 0
(VP(4,8), VERB) ukončí S ← NP VP (1,8)
S ← NP VP (2,8)

(S(1,8) S(2,8))

Analyzátor identifikoval korektnú anglickú vetu S(1,8) od prvého po ôsme slovo, a tiež S(2,8), pretože uvedená veta bez prvého slova tvorí tiež vetu. Ak chceme získať všetky alternatívne analýzy, pokračujeme, až kým nebude zoznam kľúčov prázdny.

4.7 Výhody a nevýhody analýzy zhora-nadol a zdola-nahor

Oba uvedené postupy majú svoje pozitíva aj negatíva.

Výhodou syntaktickej analýzy zhora-nadol je jej prediktívna schopnosť. Analýza sa začína symbolom pre vetu (počiatočným symbolom), preto nikdy neuvažuje gramatické kategórie v pozíciách, kde by vo vete nemohli stáť. Na základe už zanalyzovanej ľavej časti vstupného reťazca dokáže predpovedať, aké gramatické kategórie pravdepodobne budú nasledovať, ak má byť veta gramaticky správna. Napríklad ak sa slovo *can* objaví po určitom člene *the*, analyzátor ho bude automaticky pokladať za podstatné meno, nie za pomocné sloveso.

Nevýhodou postupu zhora-nadol je, že analyzátor pracuje nejaký čas predtým, než sa vôbec dostane k terminálom vstupného reťazca, ktoré mu poskytujú prvú informáciu o tom, čo sa v danej konkrétnej vete v skutočnosti nachádza. Analyzátor typicky očakáva napríklad vetu začínajúcu sa mennou frázou, ktorá sa začína substantívom atď. Ale až pri pohľade na vstupný reťazec môže zistiť, že sa začína povedzme adjektívom, a musí svoje nesprávne predpoklady prerobiť. Vykonáva teda nadbytočnú prácu. Pri procese návratu (backtracking) sa často eliminujú aj konštituenty, ktoré boli už zanalyzované správne, a pri opätovnom postupe vpred sa analyzujú ešte raz. Analyzátor v horeuvedenom príklade musel napríklad opakovať analýzu slovesnej frázy (*ťahá vozík*), pretože sprvu očakával intranzitivitu.

Pri analýze zdola-nahor sa naopak žiaden konštituent nekonštruuje dvakrát, lebo všetky sa uchovávajú v zozname konštituentov. Postup zdola-nahor tiež berie už od začiatku ohľad na skutočný konkrétny začiatok vety a jeho gramatické kategórie.

Na druhej strane je však analyzátor typu zdola-nahor odsúdený uvažovať aj všetky možné konštituenty, ktoré vyzerajú korektne, ale stoja na miestach, na ktorých nikdy nebudú môcť prispieť k legálnej vete a sú teda zbytočné. Analyzátor

z nášho príkladu zbytočne identifikoval napríklad mennú frázu uprostred vety od 2 do 4 alebo slovesnú frázu od 5 do 8. Zistil aj, že slová od druhého po ôsme tvoria vetu, čo je síce pravda, ale neužitočná, pretože prvé slovo, ktoré do vety určite patrí, by tak zostalo nezanalyzované.

	top-down (zhora-nadol)	*bottom-up (zdola-nahor)*
+	Prediktívnosť. Nikdy neuvažuje gramatické kategórie v pozíciách, kde nemôžu stáť, pretože sa začína symbolom *S*.	Nekonštruuje žiadnu frázu dvakrát.
–	Môže operovať nejaký čas predtým, než sa dostane k terminálom. Tá istá práca sa môže opakovať.	Uvažuje aj všetky frázy, ktoré nemôžu prispieť k legálnej vete.

4.8 Zmiešaná stratégia. Analyzátor zhora-nadol s listinou

Ideálny by bol, pravda, analyzátor, ktorý by v sebe zlučoval výhody oboch postupov: zhora-nadol aj zdola-nahor. Vyskytli sa pokusy o syntaktickú analýzu súčasne zhora aj zdola, avšak zlučovanie týchto dvoch takých cudzorodých stratégií, ktoré sa musia stretnúť niekde v strede syntaktického stromu, si vyžaduje komplikované kontrolné mechanizmy (Kay, 1973; Winograd, 1983).

Existuje však metóda, ktorá prirodzeným spôsobom zlučuje postup zhora-nadol s niektorými výhodami postupu zdola-nahor. Je to syntaktická analýza zhora-nadol s listinou hotových konštituentov (Allen, 1987). Vždy, keď analyzátor hodlá prepísať nejaký neterminál, pozrie sa najprv do zoznamu, či na tom mieste nie je taký konštituent už hotový. Ak áno, použije ho. Takýto analyzátor teda nevykonáva analýzu žiadnej frázy dvakrát.

SAPFO je syntaktický analyzátor zľava-doprava so zmiešanou stratégiou. Postupuje zhora-nadol s využitím zoznamu hotových konštituentov.

4.9 Gramatika Hornových klauzúl (DCG, Definite Clause Grammar)

Nepostrádateľnou súčasťou skladby je zhoda vetných členov v gramatických kategóriách. V slovenčine je to zhoda podmetu s prísudkom v rode a čísle a zhoda prívlastku s menom v rode, čísle, páde a životnosti.

Čistá bezkontextová gramatika však kategoriálne zhody ignoruje a generuje aj vety a frázy ako:

starého otcom neprišli, *vysokej gaštanoch spoza horou*

Teoreticky by sa gramatická zhoda bezkontextovým formalizmom dala vyjadriť, ale iba za cenu neprirodzeného nárastu počtu pravidiel. Napríklad pravidlo pre mennú frázu pozostávajúcu z prídavného a podstatného mena by sme museli nahradiť dvanástimi pravidlami, pre každý pád a číslo osobitne.

Tento nedostatok odstraňuje tzv. *DCG-gramatika* (Pereira - Warren, 1980). Je to vlastne bezkontextová gramatika opatrená n-ticou argumentov. Argumenty sú podriadené PROLOG-ovskej unifikácii, takže prenos gramatickej informácie sa vykonáva automaticky. Napríklad prvé a šieste pravidlo v našej gramatike by sme prepísali ako:

(1) v(Ro,Ci) → nf(Ro,Ci,_,_),sf(Ro,Ci).
(6) nf(Ro,Ci,Pa,Zi) → adj(Ro,Ci,Pa,Zi),sub(Ro,Ci,Pa,Zi).

Ku gramatickým argumentom budeme môcť neskôr pridať aj sémantické, aby gramatika generovala vety ako *Janko plače* ale nie *Kameň plače* alebo *„Janko plače slzy.*

(1) v(Ro,Ci,Živ) → nf(Ro,Ci,_,Živ),sf(Ro,Ci,Živ).
(3) sf(Ro,Ci,Živ) → sloveso(intranz,Ro,Ci).
(4) sf(Ro,Ci,Živ) → sloveso(tranz,Ro,Ci),nf(_,_,akuz,_).

Ďalšie rozšírenie je vynútené skutočnosťou, že potrebujeme nielen *rozhodnúť*, či daná veta patrí do jazyka alebo nie, ale potrebujeme *skonštruovať* nejakú štruktúru (syntaktický strom, rámec...), ktorá bude predstavovať výsledok analýzy. Na pravých stranách pravidiel pribudnú tzv. *konštrukčné návestia*. V konštrukčnom návestí je zoznam akcií, ktoré interpreter vykoná, ak pri vykonávaní pravidla na takéto návestie narazí.

(1) v(Ro,Ci,Živ) → nf(Ro,Ci,Živ),sf(Ro,Ci,Živ),<*cons SUBJEKT*>.
(3) sf(Ro,Ci,_) → sloveso(Ro,Ci),nf(_,_,akuz,_),<*cons OBJEKT*>.

Výsledkom analýzy je potom hniezdovaná štruktúra, ktorá jednotlivým konštituentom priraďuje ich syntaktické roly:

PREDIKÁT: ťahať
SUBJEKT: Jano
OBJEKT: [vozík
  PRÍVLASTOK: malý]
PRÍSLOVKOVÉ URČENIE: les

4.10 Čo robí dobrú gramatiku dobrou gramatikou

- *Všeobecnosť*. Množina gramatických viet, ktoré gramatika dokáže akceptovať.

- *Selektivita*. Množina negramatických viet, ktoré gramatika dokáže vylúčiť.

- *Elegancia, jednoduchosť*. Táto požiadavka je zdanlivo iba estetická, ale pri veľkých gramatikách so stovkami až tisíckami pravidiel sa ukáže, že je rozhodujúca.

4.11 Reverzibilné gramatiky

Inou žiadúcou charakteristikou dobrej gramatiky je *reverzibilnosť*. Pravidlá reverzibilnej gramatiky sa dajú používať v oboch smeroch - sprava-doľava aj zľava-doprava. To umožňuje použiť jednu a tú istú gramatiku pre syntézu aj analýzu.

SAPFO je čiastočne reverzibilná sémantická gramatika. Valenčné a kolokačné funkcie (VAL, LVAL, RVAL, CVAL, DVAL) sú reverzibilné a pri analýze a generovaní sa používajú obojsmerne (kap. 5,7,12).

4.12 Prechodová sieť (TN, Transition Network)

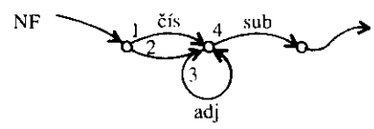
Pomocou DCG-gramatiky možno už generovať pekné a dosť senzitívne podmnožiny prirodzeného jazyka. Ak si však postavíme za cieľ zachytiť slovenskú skladbu v jej celistvosti, t. j. napísať *všetky*, nielen *niektoré* pravidlá slovenskej syntaxe, ukáže sa, že to nie je možné.

Uvažujme, že by sme chceli vypísať pravidlá pre nominálnu frázu. Na pravej strane môže byť menná fráza rozvitá genitívnym alebo datívovým prívlastkom, radom predložkových fráz, infinitívom, adjektiválnou frázou, ďalšou mennou frázou, prístavkom, vedľajšou vetou. Na ľavej strane môže stáť záporová častica, základné, druhové, radové, násobné, zlomkové a skupinové číslovky, vymedzovacie, odkazovacie a privlastňovacie zámená, hodnotiace a vzťahové adjektíva. Z týchto prvkov sú niektoré povinné, niektoré nepovinné, niektoré môžu byť viacnásobné a všetky sa kombinujú v určitom, ale nie ľubovoľnom a nie zameniteľnom poradí (pozri sieť na obr. 4). Dostávame 147 456 možných kombinácií či pravidiel pre mennú frázu!

Povedzme, že by sme uvažovali iba menné frázy s jednou nepovinnou číslovkou a nula až trojnásobným prídavným menom. Vyžadovalo by to 8 pravidiel:

nf → sub.	nf → čís,sub.
nf → adj,sub.	nf → čís,adj,sub.
nf → adj,adj,sub.	nf → čís,adj,adj,sub.
nf → adj,adj,adj,sub.	nf → čís,adj,adj,adj,sub.

To isté sa dá vyjadriť *prechodovou sieťou* so štyrmi hranami. Horeuvedená osmica pravidiel je čo do generatívnej sily ekvivalentná s takouto sieťou:



Prechodová sieť je graf, ktorého hrany sú označené slovnodruhovými a gramatickými kategóriami. Frázy sa generujú alebo analyzujú tak, že cez sieť prechádzame rozličnými cestami. Napríklad menná fráza *štyri biele vrany* vznikne po prechode hranami 1,3 a 4. Fráza *vrana* vzniká prechodom hranami 2 a 4 atď.

Prechodová sieť elegantne rieši opakovanie a nepovinnosť. Viacnásobné vetné členy sú vyjadrené slučkou (3) a fakultatívne členy prázdnymi, neoznačenými hranami (2, jump arcs). Rekurzívna prechodová sieť (RTN) dovoľuje volať aj podsiete.

4.13 Rozšírená prechodová sieť (Augmented Transition Network)

Rozšírené prechodové siete (ATN) zaviedol W. Woods (1969) v dialógovom systéme LUNAR, určenom na analýzu vzoriek mesačných hornín.

Predstavme si, že máme v DCG gramatike (okrem iných) pravidlo:

veta → nf(Ro,Či,Pa,Ži,Os),adverb,verb_fin(Či,Ži,Os).
veta → ...

a analyzujeme vetu:

Môj mladší brat Filip tam išiel pomôcť.

Analyzátor vyberie prvé pravidlo a správne odhalí na začiatku vety nominálnu frázu *môj mladší brat Filip*, potom príslovku *tam* a určité sloveso *išiel*. Teraz sa ukáže, že na konci vstupného reťazca zostalo ešte jedno slovo (*pomôcť*), teda pravidlo bolo vybrané nesprávne; pokus o analýzu treba obnoviť vybraním nasledujúceho pravidla pre vetu. Prečo však zahodiť hotovú analýzu celej prvej časti vety (až na posledné slovo), ktorá bola správna ? Túto neefektivitu odstraňuje ATN-sieť.

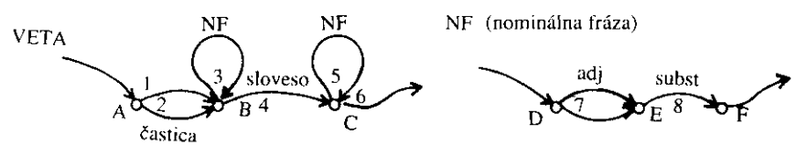
Rozšírená prechodová sieť je syntaktický analyzátor pracujúci na princípoch „zľava doprava" a „zhora nadol". Pozostáva z orientovaného *grafu*, zoznamu *podmienok* a zoznamu *akcií*. Vrcholy grafu zodpovedajú stavom automatu, do ktorých sa analyzátor môže dostať. Ku každej hrane je priradená podmienka, pri ktorej možno cez hranu prejsť, a akcia, ktorú treba pritom vykonať.

Ak analyzátor postúpi vo vstupnej vete o jedno slovo, postúpi zároveň v sieti cez jednu hranu. Pritom priebežne konštruuje výsledný syntaktický strom. Ak je z daného stavu viac možných spôsobov postupu, vyberie jeden a ostatné uloží do zásobníka. Ak sa analyzátor ocitne v „slepej uličke", tzn. podmienky nedovoľujú

v danom stave traverz po žiadnej výstupnej hrane, vykoná sa proces návratu. Zo zásobníka sa vyberie stav, v ktorom mal analyzátor posledne na výber a pokračuje sa iným spôsobom.

Ak je zásobník prázdny, nie je možné postúpiť a hlava automatu nie je na konci vstupného reťazca - veta nepatrí do daného jazyka. Ak sa analyzátor dostane zároveň na koniec siete aj na koniec vstupnej vety, analýza bola úspešná a výsledok analýzy je reprezentovaný v príslušnej stromovej (rámcovej) štruktúre.

Ilustrujme prácu syntaktického analyzátora na malej ATN-sieti ('*' označuje práve analyzovaný vstupný symbol):



Hrana	Podmienky	Akcie
2.	* = častica(_)	
3.,5.	* = nf(_,Pa,Ci,Ro,Zi)	ak Pa=nom CIS ← Ci, ROD ← Ro, ZIV ← Zi, SUBJEKT ← * ak Pa=akuz OBJEKT ← * ak Pa=dat NEPRIAMY_OBJEKT ← *
4.	* = sloveso(_,_,_,ROD,CIS,ZIV)	PREDIKÁT ← *
6.	čítacia hlava na konci vstupnej vety	
7.	* = adjektívum(_,P1,C1,R1,Z1)	DETERMINANT ← *
8.	* = substantívum(_,P1,C1,R1,Z1)	

Podmienky k hranám 3, 4, 5, 7, 8 požadujú zhodu vetných členov v rode, čísle a páde. Akcie na hranách 3, 4, 5 a 7 priraďujú syntaktické roly, ktoré budú výsledkom analýzy.

Stav automatu je daný obsahom zásobníka a dvojicou:

(vrchol ATN-siete, poloha čítacej hlavy vo vstupnej vete).

Napr. (C,4) znamená: nachádzame sa vo vrchole C (siete) a čítacia hlava ukazuje za štvrté slovo vstupnej vety.

Analyzujme (trochu umelú) vetu:

$_0$ ***Veď*** $_1$ ***kvapkám*** $_2$ ***slnko*** $_3$ ***zlatí*** $_4$ ***okraje*** $_5$

Morfologická analýza nám dala tento výstup:

veď	→	sloveso(viesť,rozkaz,2.os,sg).
veď	→	častica(uvádzacia).
kvapkám	→	sloveso(kvapkať,prít_čas,1.os,sg).
kvapkám	→	substantívum(kvapka,dat,pl,muž,neživ).
slnko	→	substantívum(slnko,akuz,sg,str,neživ).
slnko	→	substantívum(slnko,nom,sg,str,neživ).
zlatí	→	adjektívum(zlatý,nom,pl,muž,živ).
zlatí	→	sloveso(zlatiť,prít_čas,3.os,sg).
okraje	→	substantívum(okraj,nom,pl,muž,neživ).
okraje	→	substantívum(okraj,akuz,pl,muž,neživ).

Začiatok analýzy. Nachádzame sa vo vrchole A, pred prvým slovom vstupnej vety; zásobník je prázdny:

(A,0) Z = 0

Postúpime po hrane č. 1 (bez podmienky) do vrcholu B. Možnosť prejsť s časticou *veď* cez hranu č. 2 si uložíme do zásobníka.

(B,0) Z = (B,1)

Vo vrchole B sieť očakáva nominálnu frázu - volanie podsiete pre *nf* však zlyhá - tvar *veď* nie je ani prídavným, ani podstatným menom. Prejdeme teda cez hranu 4 a identifikujeme *veď* ako prísudok vety:

(C,1) Z = (B,1) PREDIKÁT = viesť

Vo vrchole C sieť očakáva rad nominálnych fráz (nf), ktoré naozaj nájde - *kvapkám* a *slnko:*

(C,3) Z = (B,1) PREDIKÁT = viesť
OBJEKT = slnko
NEPRIAMY_OBJEKT = kvapky

Opustiť sieť po hrane č. 6 nie je možné, lebo sme sa ešte nedostali na koniec vety. Zostáva len nájsť ďalšiu NF. Nasledujú slová *zlatí* a *okraje*, ktoré naozaj predstavujú prídavné meno nasledované podstatným menom; netvoria však frázu, lebo sa nezhodujú v životnosti. ATN sa ocitla v slepej uličke. Musí teda vybrať zo zásobníka stav (B,1), t. j. vrátiť sa na začiatok vety a reinterpretovať tvar *veď* ako časticu, nie predikát:

(B,1) Z = 0

Nasleduje tvar *kvapkám*. Sieť očakáva NF (túto možnosť si odložíme do zásobníka) alebo sloveso - môže prejsť po 4. hrane:

(C,2) Z = (B,2) PREDIKÁT = kvapkať

ATN-sieť sa ešte raz snaží nájsť rad NF *slnko* a *zlatí okraje*; a neúspech ju znova donúti k návratu:

(B,2) Z = 0

Konečne sa analýza dostáva na správnu cestu. Sieť odhalí *kvapkám* a *slnko* ako nominálne frázy v datíve a nominatíve a priradí im príslušné roly:

(B,3) Z = 0 SUBJEKT = slnko
NEPRIAMY_OBJEKT = kvapky

ďalej môže prejsť (len) cez 4. hranu so *zlatí* ako slovesom:

(C,4) Z = 0 SUBJEKT = slnko
NEPRIAMY_OBJEKT = kvapky
PREDIKÁT = zlatiť

Zostávajú *okraje* ako NF v akuzatíve:

(C,5) Z = 0 OBJEKT = okraje

Pretože sme šťastne dosiahli koniec vstupného reťazca, môžeme ATN-sieť opustiť; veta je z jazyka generovaného touto sieťou a výsledná analýza vyzerá takto:

PREDIKÁT = zlatiť
SUBJEKT = slnko
OBJEKT = okraje
NEPRIAMY_OBJEKT = kvapky

4.14 Porovnanie generatívnej kapacity sietí a gramatík

Ak by sme chceli porovnať generatívnu silu gramatík a sietí, tak regulárne, bezkontextové a kontextové gramatiky by *zhruba* zodpovedali jednoduchým, rekurzívnym a rozšíreným prechodovým sieťam. Avšak prechodová sieť, komentovaná špecifickými podmienkami a akciami pri každej hrane, umožňuje vytvoriť oveľa efektívnejší a senzitívnejší aparát

gramatiky	siete
REG	TN
CFG	RTN
CSG	ATN

Vychádzajúc z Winogradových (1983) a Allenových (1987) myšlienok, pokúsili sme sa využiť techniku prechodových sietí aj pre slovenský jazyk. Otázka znela: je možné aplikovať techniku rozšírených prechodových sietí na flektívne jazyky, a ak áno, aké adaptácie si takáto aplikácia vyžaduje ?

4.15 Deterministickosť

Opakované prehľadávanie a proces návratu zo slepých uličiek intuitívne nezodpovedá činnosti, ktorú vykonáva pri analýze jazyka človek. Človek dokáže vnímať vetu spravidla bez toho, aby sa musel vracať na začiatok. Žiadúca je teda skorodeterministická analýza.

Za účelom včasnej eliminácie nejednoznačností už počas analýzy boli navrhované rozličné heuristické pravidlá, ako *pravidlo minimálneho počtu závislostí* alebo *pravidlo zružovania napravo* (Allen, 1987). Všetky sú však nedostatočné. M. Marcus (1980) navrhol analyzátor, ktorého *predhľad* (lookahead) je väčší ako jedno (dve až tri) slová dopredu. Navrhovali sa aj *orákulá*, automaty, ktoré sú obrazom všetkých možných stavov analyzátora, aké sa pri analýze vety môžu vyskytnúť. Ďalšia možnosť je zaviesť *lexikálne preferencie*.

Podľa nášho názoru človek dokáže nezmyselné syntaktické stromy okamžite vylučovať preto, lebo si uvedomuje lexikálnu sémantiku slov už počas syntaktickej anlýzy. Preto riešenie tohto problému vidíme iba v plnej paralelizácii, úzkom prepojení procesov syntaktickej a sémantickej analýzy. Ako uvidíme v kapitole 6, podarilo sa nám k želanému sub-deterministickému analyzátoru značne priblížiť.

4.16 Rozšírené prechodové siete (ATN) pre slovenský jazyk

Výsledkom experimentov so systémom SAPFO sú rozšírené prechodové siete pre slovenský jazyk (Páleš, 1990). Vznikli iteratívnym ladením metódou pokusov a omylov na vzorke 1000 slovenských viet rozličných štýlov. Sieť pozostáva

z ôsmich podsietí, ktoré navzájom komunikujú a môžu rekurzívne volať jedna druhú: *súvetie, veta, nominálna fráza, predložková fráza, adjektiválna fráza a participant*. Podsiete pre *vedľajšiu vetu a infinitívnu konštrukciu* sa vlievajú do siete pre vetu.

Kostra grafu so slovnodruhovými kategóriami je znázornená na obrázkoch 3 až 8. Ostatné podmienky a akcie k sieti sú stelesnené v procedúre programu, a to hlavne v module syntaxe (pozri listing programu - moduly VI, VII). Sémantické podmienky k sieti sú dané kolokačnými rovnicami VAL, LVAL, RVAL (moduly VIII, IX a XVI).

Syntaktický analyzátor SAPFO je teda ATN-sieť, ktorá využíva unifikáciu Hornových klauzúl podobným spôsobom ako DCG-gramatika. Je to druh unifikačnej gramatiky (Allen, 1987) so silným sémantickým komponentom.

Uvedené siete obsiahnu značné množstvo javov slovenskej syntaxe: konštrukcie aktívne aj pasívne, infinitívne, prechodníkové, príčasťové a polopredikačné (prístavok, doplnok), vedľajšie vety a súvetia; nezabudli sme ani na syntax čísloviek, miesta pre častice a citoslovcia a osobitné parentetické útvary oddelené interpunkčnými znamienkami.

Napriek tomu ešte stále zostáva nepreberná trieda javov, ktoré analyzátor neakceptuje. Neuvažovali sme napríklad výskyt absolútneho nominatívu či akuzatívu (*Dedko sedí na lavičke,* ***fajka v ústach****, zdá sa, že nevníma)*, ani doplnky umiestnené inde ako na konci vety (***Zamyslená*** *sedela Kassandra na morskom brehu a dívala sa do diaľky*). Analyzátor nie je pripravený na deformovanú syntax hovorenej reči. Osobitným problémom sú koordinácie a elpisy (4.21).

4.17 Porovnanie slovenských a anglických ATN

Slovenské prechodové siete majú v porovnaní s anglickými komplexnejšiu štruktúru. Voľný slovosled slovenčiny sa odráža vo väčšom počte hrán. Porovnajme len slovenskú NF (obr. 4) s elegantnou jednoduchosťou vzorca pre anglickú NP (Winograd, 1977):

NP = <determiner>,{ordinal},{number},{adjective}*,{classifier}*,
<head>,{qualifier}*.

Táto komplexnosť slovenskej siete je však na druhej strane kompenzovaná jej väčšou deterministickosťou. Pre anglickú syntaktickú analýzu je charakteristický proces návratu. Slovnodruhová amorfnosť anglických slov (napr. *love* je zároveň slovesom, podstatným aj prídavným menom) často zavádza analyzátor do slepých uličiek; už hotové časti analýzy sa musia aj niekoľkokrát prepracovať. Anglická morfológia je zredukovaná na reziduálne formy na úkor syntaxe, ktorá je extrémne zaťažená. Teda analyzátor sa síce pohybuje po menšej sieti, ale po mnohých hranách prejde viackrát.

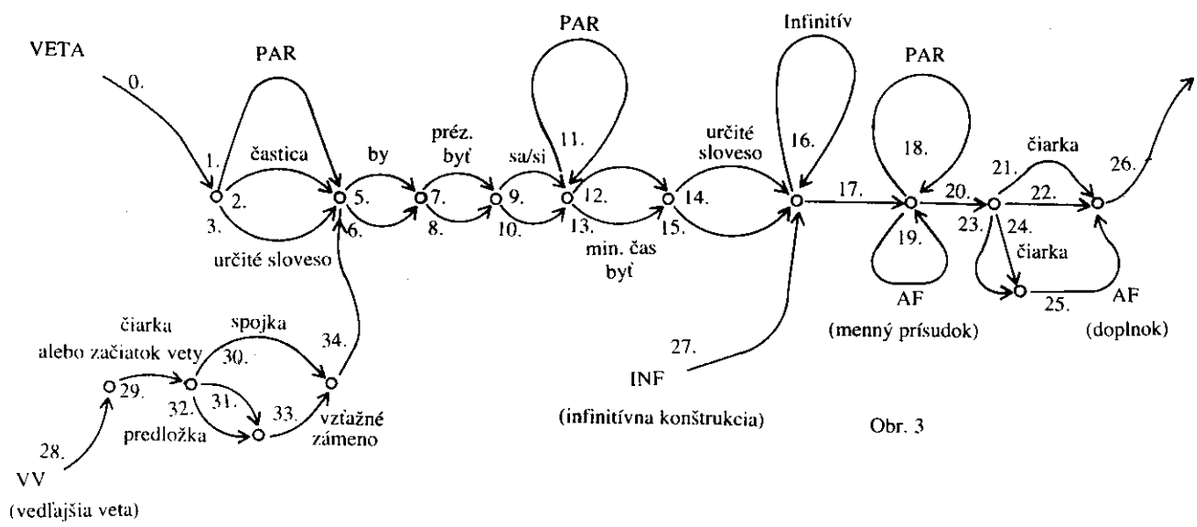
Zradný príklad 4.13 je pre slovenčinu netypický, pretože obsahuje veľa tvarovej homonymie. Slovenský analyzátor postupuje viac deterministicky, opierajúc sa

ATN

(Rozšírené prechodové siete)
pre
slovenský jazyk

PAR PAR Infinitív PAR VETA

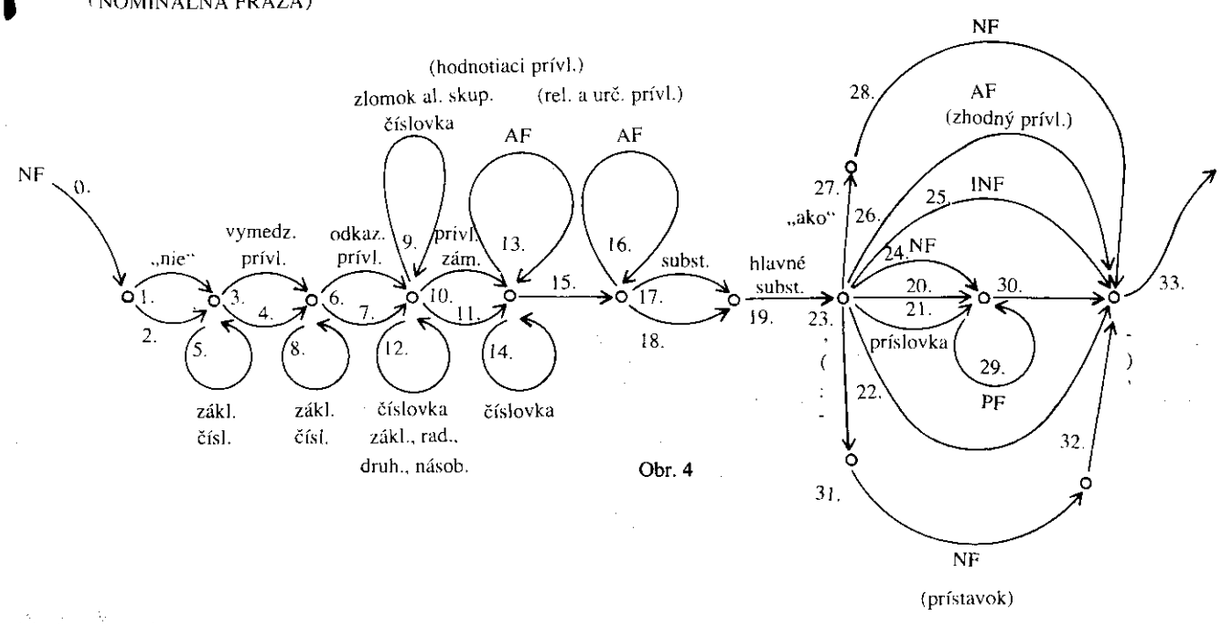
Veta



Obr. 3

Obr. 3: ATN

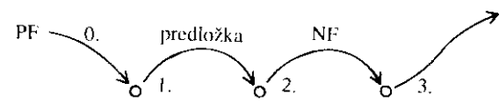
nf (NOMINÁLNA FRÁZA)



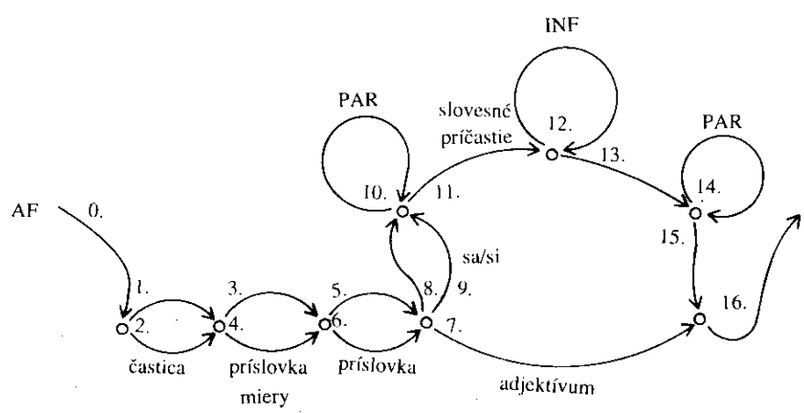
Obr. 4

Obr. 4:

PF (PREDLOŽKOVÁ FRÁZA)

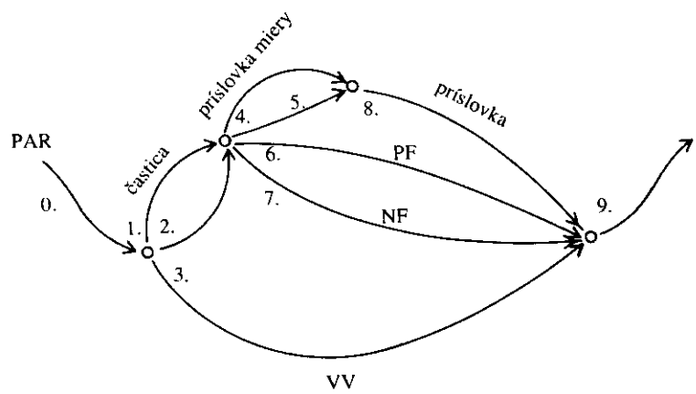


AF (ADJEKTIVÁLNA FRÁZA)

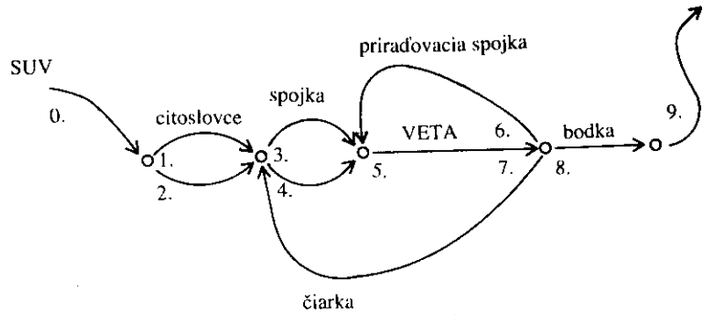


Obr. 5

PAR (PARTICIPANT)



SUV (SÚVETIE)



Obr. 6

o bohatý pádový systém flektívneho jazyka. Slovnodruhová príslušnosť slov ako aj syntaktické roly aktantov sú zviditeľnené morfologicky a analyzátor ich zväčša určí správne už na prvý raz.

Bohatosť tvaroslovnej štruktúry teda nepokladáme za komplikáciu či prekážku, ale naopak za skutočnú výhodu flektívnych jazykov. Zdanlivo nadbytočné tvarové bohatstvo veľmi odľahčuje syntax. Rovnomerná vyváženosť skladby, tvaroslovia a slovotvorby je z hľadiska výpočtovej zložitosti ideálna (pozri 16.5).

Vcelku sa dá povedať, že SAPFO je implementáciou a zároveň potvrdením hypotézy, ktorú vyslovil T. Winograd (1983, s. 324): „...in English, not much has been done to relate computational uses of (semantic) case to the phenomena of surface case. In a system for analyzing a language with more developed case system, like Russian or Japanese, the use of verb-centered case analysis would be even of more benefit. In these languages the parser cannot rely on the order of constituents, looking for a sequence like 'SUBJECT VERB OBJECT'. It must operate by picking up the components of the sentence and fitting them into case roles, using the surface case information to decide which constituent fills which role."

4.18 Oslabenie prediktívnej sily

Technika ATN bola vyvinutá na experimentálnom materiáli angličtiny. Jej hlavná sila spočíva v prediktívnej schopnosti, ktorá sa opiera o pevný slovosled analytických jazykov. Anglický syntaktický analyzátor môže s veľkou pravdepodobnosťou predpokladať vetné členy v poradí *subjekt - predikát - nepriamy objekt - priamy objekt - predložková fráza - príslovka.*

Voľný slovosled slovenčiny však neumožňuje na základe analýzy začiatku vety predpokladať, aký vetný člen bude nasledovať. Prílišná prediktívna sila sa tu stáva nevýhodou.

Hoci sa nám podarilo prediktívnu schopnosť slovenských sietí na príslušných miestach vhodne oslabiť (napríklad sieť pre vetu neočakáva subjekt, objekt a pod. na určitých miestach, ale len aktanty, voľne sa vyskytujúce okolo slovesa), istá neefektívnosť je tu stále zreteľná. Napríklad typická nominálna fráza bude v 80 %-tách prípadov pozostávať z holého podstatného mena, prípadne rozvitého jedným členom sprava alebo zľava. Avšak sieť pre NF (obr. 4) je nútená vždy hľadať a skontrolovať výskyt všetkých druhov atribútov - jedného po druhom (hrany 3, 5, 6, 8, 9, 10, 12, 13, 16...), aby nakoniec našla iba jeden alebo žiaden z nich.

Vo všeobecnosti sa ATN-siete pre syntaktickú analýzu slovenčiny dajú použiť; i keď s istou stratou efektivity. **Pre jazyky flektívneho typu by sme potrebovali skôr mechanizmus, ktorý nie je predikčný; t. j. nič vopred nepredpokladá, ale radšej sa najprv pozrie na začiatok vstupného reťazca a podľa toho koná.**

Podľa toho, čo sme povedali v odseku 4.7, tento výsledok hovorí skôr v prospech použitia metód syntaktickej analýzy **zdola-nahor**, ktoré by mali byť **pre jazyky s voľným slovosledom vhodnejšie**, ako postup zhora-nadol.

4.19 (Polo)voľný slovosled slovenčiny

Avšak slovenský slovosled zďaleka nie je taký „voľný", ako by to tento zaužívaný termín naznačoval. Slovesná zložka slovenskej vety tvorí viac-menej pevnú kostru. Pozostáva z pomocných častíc a slovies, určitého slovesa a neurčitku (hrany 5, 7, 9, 13, 14, 16 na obr. 3). Prvky tejto skupiny sú síce fakultatívne, ale ak sú prítomné, neslobodno ich preradiť alebo medzi ne vkladať. Na tejto kostre je niekoľko miest pre aktanty (menné a predložkové frázy, vedľajšie vety, príslovkové určenia), ktoré obiehajú vo voľnom poradí. Voľbou poradia sa vyjadruje aktuálne členenie.

Ľavá strana nominálnej frázy tiež vykazuje viac pevný než voľný slovosled. Zoradenie zhodných prívlastkov v poradí

vymedzovací-odkazovací-privlastňovací-hodnotiaci-relačný

je záväzné:

všetky pekné domy	pekný červený kvet
*pekné všetky domy	*červený pekný kvet

Určité porušenie štandardného poradia prívlastkov je možné, ale je vždy štylisticky príznačné. Vzniká zaujímavá otázka: ak zmenou poradia vo vete vyjadrujeme aktuálne členenie, prečo sme si ponechali možnosť prehadzovať poradie čísloviek a iných prívlastkov v nominálnych frázach?

4.20 Polopredikácie vnútri nominálnych fráz

V záujme kondenzácie (zhustenia) prejavu vety často obsahujú okrem hlavného predikátu (vyjadreného verbom finitom) súčasne aj ďalšie, vedľajšie tvrdenia, tzv. *polopredikácie*. Ako polopredikatívne konštrukcie sa obyčajne uvádzajú len prechodník, prístavok a doplnok.

Chceme tu poukázať na fakt, že polopredikácie sa tvoria aj vnútri nominálnych fráz, a že tomuto účelu je vyhradená práve inverzia slovosledu. Všimnime si:

Dve moje dcéry sa vydali. (= Dve z mojich dcér sa vydali.)

Prívlastky sú v štandardnom poradí a majú reštrikčnú funkciu. Po inverzii číslovky sa objaví druhé tvrdenie:

Moje dve dcéry sa vydali. (= 1. Moje dcéry sa vydali.
2. Moje dcéry sú dve.)

Podobné zmeny badať aj pri presune iných prívlastkov z antepozície do postpozície:

moje zlaté dcéry	→	dcéry (vy ste) moje zlaté !

Nemyslíme, že ide o náhodný jav. Napríklad v taliančine je tvorenie polopredikácií inverziou gramatikalizované. Základné postavenie talianskeho prívlastku je v postpozícii. Pred podstatným menom znamená polopredikáciu:

la ragazza bella	=	pekné dievča
bella ragazza	=	(to je) pekné dievča (!)

Na tento účel je ATN-sieť vyzbrojená aj hranami pre vetné členy a prívlastky v neštandardných syntaktických pozíciách. K takýmto hranám je pripojený dodatočný zoznam zvláštnych akcií, ktoré treba vykonať.

Napríklad hrana č. 14 v sieti NF slúži na zachytenie čísloviek v inverznej pozícii. Ak sa číslovka objaví až na tejto hrane, ide o inverziu a sieť automaticky konštruuje a pripája k výslednej analýze ešte jedno postranné tvrdenie - polopredikáciu. (Pozri aj 10.6 - štylistická hodnota syntaktických konštrukcií.)

4.21 Koordinácie a elipsy

SAPFO počíta s niektorými najčastejšími druhmi priraďovacích skladov na úrovni viet a menných fráz. Ale konjunkcie, disjunkcie a elipsy (výpustky) sa môžu vyskytnúť na úrovni každého vetného člena:

koordinácie	**na úrovni**
Padal včera sneh alebo zasa (včera) iba pršalo ?	vety
Škaredý Bill a pekná Mary.	nom. fr.
Pájka sa vyrába zo zlata a z ľahko taviteľného kadmia.	predl. fr.
Ťažko sa nájde taký priestranný a slnkom presvetlený byt.	adj. fr.
Janko, Marienka a ježibaba vošli do chalúpky.	podst. m.
Matka ju umyla a učesala.	slovesá
Prvé, druhé a tretie miesto budú odmenené.	rad. čísl.
Dva či tri vrabce sadli na konár.	zákl. čís.
Uvážil všetky argumenty za a proti tejto téze.	predložky

elipsy

Janko má rád jahodovú zmrzlinu a Marienka (má rada) vanilkovú (zmrzlinu).
Tebe sa to podarilo a mne (sa to) nie. (podarilo)
Kedy asi príde ? Myslím, že (príde) v piatok.

Koordinácie a elipsy už nepredstavujú iba technickú, ale principiálnu prekážku sieťovej analýzy.

Zahrnúť do prechodovej siete koordinácie by vyžadovalo prakticky zdvojnásobiť počet hrán, pretože každý vetný člen môže byť viacnásobný. A elipsovité (neúplné) vety sú pre techniku ATN tak cudzorodým prvkom, že snaha o ich začlenenie do siete by vyústila do popretia všetkého, čo sme doteraz urobili. V eliptickej vete môže byť jednoducho vynechaný ktorýkoľvek vetný člen, hlavný či vedľajší, viac členov, alebo aj všetky okrem jedného. Podľa toho by sme museli prakticky každú hranu v prechodovej sieti vyhlásiť za nepovinnú, čím by vlastne ATN-analýza bola anulovaná. Prechodová sieť, silná svojimi predikčnými schopnosťami typu „za tým a tým musí nasledovať to a to" by všetku svoju silu stratila - v elipsovitej vete nemusí nasledovať nič.

Výpovede s elipsami nás upozorňujú na jeden fundamentálny jav, ktorý sa vymyká formalizácii: *Účastníci komunikácie jazykom vyjadrujú, vlastne iba doplňujú vždy len to, čo už nie je zrejmé z ich spoločného mimojazykového pozadia.* Úplné vety sa v dialógu ani nesnažíme formulovať, kým to nie je nevyhnutné.

Rekonštrukcia koordinácií s elipsami. Ak sa determinatívna syntagma stane súčasťou koordinatívnej, vzniká typický problém: ako jednoznačne určiť dosah determinatívnej syntagmy ? Vo fráze

červené ruže a tulipány

sa prívlastok *červené* môže vzťahovať iba na ruže alebo na ruže aj tulipány. Môže ísť o výpustku, zjednodušenie frázy *červené ruže a červené tulipány*.

SAPFO narába s determináciami v koordináciách tak, že širší dosah determinantu predpokladá iba vtedy, ak je fráza syntakticky neúplná a rekonštrukciu pôvodných členov si vynucuje:

pekné červené a žlté tulipány

Eliptické vety dokážeme formálne rekonštruovať iba vtedy, keď sa nachádzajú súčasne v koordináciách. Prvá časť priraďovacieho skladu, ktorá je v plnom znení, poskytuje šablónu, na základe ktorej sa v druhej časti skladu doplní všetko, čo tam chýba. Prázdne syntaktické miesta sa doplnia tými istými plnovýznamovými slovami, ako v prvej časti skladu:

Janko má rád jahodovú zmrzlinu a Marienka má rada vanilkovú.

Druhá veta súvetia vyžaduje, aby bola syntakticky zúplnená. Abstraktné vzorce týchto viet kontrastujú:

1. sub-sl-adj-sub.
2. sub-sl-adj-....

Z prvej vety teda preberieme kontrastné substantívum a pridáme ho do druhej (*Marienka má rada vanilkovú* ***zmrzlinu***).

Iné druhy výpustiek, ako v koordináciách, SAPFO riešiť nedokáže. Neporadí si napríklad so *situačnými elipsami*:

Vrchný, (prineste mi) minerálku!
Ja (hovorím) o koze a ty o voze!

Rekonštrukcia situačnej elipsy vyžaduje náročnú inferenciu nad komplexnou bázou poznatkov o situačných typoch (rámcoch).

4.22 Taxonomická vzdialenosť pri riešení výpustiek

Pri rekonštrukcii výpustiek teda *vstupný fragment* vždy zodpovedá nejakému *cieľovému fragmentu* v predchádzajúcej vete:

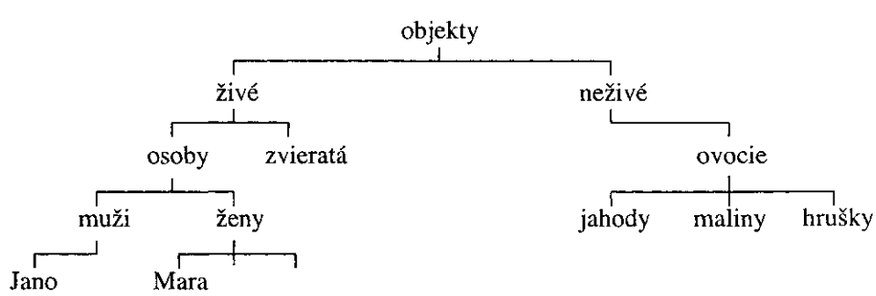
- *Dal si maslo do chladničky?*
- *Áno.*
- *A chlieb do košíka?*

Alebo ho dopĺňa:

- *Po koľko sú banány?*
- *Po dvadsať.*
- *Za kilo?*

Ak je vstupný fragment príliš krátky a z hľadiska formy by mohol zodpovedať viacerým cieľovým fragmentom v pôvodnej vete, možno využiť nasledovnú metódu. Určíme *vzdialenosť đ* medzi vstupným a cieľovým fragmentom *v taxonomickej hierarchii pojmov* a zvolíme cieľový fragment s minimálnou *đ*:

- *Videl si Jana jesť jahody?*
- *Nie.*
- *Maliny?* — *(= Videl si ho jesť* ***maliny****?)*
- *Marienku?* — *(= Videl si* ***ju*** *jesť jahody?)*



δ(Marienka,Jano) = 4
δ(Marienka,jahody) = 7
δ(maliny,Jano) = 7
δ(maliny,jahody) = 2

min δ(Marienka,X) = 4 pri X = Jano
min δ(maliny,X) = 2 pri X = jahody

Maliny teda zodpovedajú cieľovému fragmentu *jahody* a *Marienka* zodpovedá *Janovi*.

Ani táto heuristická metóda však nefunguje vždy:

- *Videl si Jana jesť medovníky?*
- *Nie.*
- *Ježibabu?*

5.

SÉMANTICKÁ ANALÝZA

5.1 Sémantická štruktúra vety

Naše poňatie vety z hľadiska významovej stavby je nasledovné: sloveso vyjadrujúce dej (činnosť) považujeme za centrum vety. Okolo neho sú potom združené jednotlivé participanty (aj aktanty), ktoré sa na deji rôznym spôsobom zúčastňujú (participujú). Sú to tie vetné členy, ktoré prichádzajú do styku s verbom finitom. Participanty môžu byť vyjadrené mennými frázami, predložkovými frázami, príslovkami, adjektiválnymi frázami, infinitívom a vedľajšími vetami.

Každý participant môže byť ďalej štruktúrovaný obdobným spôsobom. Napríklad nominálna fráza sa skladá z centrálneho mena, okolo ktorého cirkulujú participanty vo forme zhodných a nezhodných prívlastkov.

Z hľadiska sémantiky každý participant vystupuje vo vzťahu k hlavnému členu vety (frázy) v určitej sémantickej role (páde). *Vykonať sémantickú analýzu znamená identifikovať jednotlivé významové participanty vo vete a priradiť im príslušnú sémantickú rolu.*

5.2 Rozdiel medzi syntaktickými a sémantickými rolami

Zatiaľ čo syntaktická analýza priraďuje jednotlivým aktantom (frázam a podfrázam) ich syntaktické roly (podmet, predmet, prívlastok...), sémantická analýza priraďuje roly sémantické (agens, paciens, inštrument...). Medzi týmito dvoma rovinami nie je jedno-jednoznačné priradenie. Jeden sémantický pád možno vyjadriť viacerými gramatickými pádmi a jeden gramatický pád môže vyjadrovať viacero sémantických pádov.

Napríklad vo vetách (1) - (5) vystupujú okolo centrálneho slovesa *otvoriť* tri hlavné participanty - *Karol, dvere, kľúč.*

(1) Karol otvoril dvere kľúčom.
(2) Pomocou kľúča sa Karolovi podarilo otvoriť dvere.
(3) Kľúč v Karolových rukách zľahka otvoril dvere.
(4) Dvere boli otvorené Karolom.
(5) Dvere boli otvorené kľúčom.

Napriek syntaktickým odlišnostiam uvedených viet vystupuje *Karol* vždy v role AGENSa - aktívneho pôvodcu činnosti, ktorou zasahuje pasívny PACIENS - *dvere. Kľúč* má v (1) - (5) úlohu nástroja (INŠTRUMENTu), ktorým sa činnosť vykonáva. T. j. vo všetkých prípadoch ide o tých istých participantov deja v tých istých významových vzťahoch:

otvoriť
AGS» Karol
PAC» dvere
INS» kľúč

Na druhej strane vety (6), (7) majú totožnú povrchovú, ale rozdielnu významovú štruktúru:

(6) *Odrezal som Karolovi chlieb.* (7) *Odrezal som Karolovi ruku.*

odrezať	odrezať
PAC» chlieb	PAC» ruka
BEN» Karol	PAR» Karol

Roly jednotlivých participantov sú teda nezávislé od povrchovej štruktúry vety, preto sa sémantickým pádom často hovorí aj *hĺbkové pády*. Hĺbkový pád je charakteristický svojím jedno-jednoznačným vzťahom k významu (pozri 0.15). Táto jeho vlastnosť je dôležitá pri inferencii a tvorbe parafráz. Počítač nesmie priradiť rozličným participantom rovnakú rolu (napr. *kľúču* a *Karolovi* vo vetách (4), (5)), lebo ich bude nesprávne zamieňať. Na druhej strane Karolovi vo vetách (1) - (5) nesmie priradiť odlišné roly, aj keď vystupuje v rozličných gramatických pádoch.

5.3 Sémantické pády

SAPFO pracuje s množinou 66-tich sémantických pádov v doluuvedenej tabuľke. Pri sémantickej analýze nastáva *faktorizácia, diskretizácia* významového kontinua do 66 škatuliek; každý vzťah v reálnom svete sa musí zobraziť na jeden zo 66-tich abstraktných typov vzťahov.

Detailný výklad a vymedzenie sémantických pádov sme uviedli v Jazykovednom časopise (Páleš, 1990).

NÁVRH SYSTÉMU SÉMANTICKÝCH PÁDOV PRE SLOVENSKÝ JAZYK		
Materiálne procesy		
AGENS	AGS	(životný pôvodca cieľavedomej činnosti)
PROCESOR	PCS	(entita, ktorej činnosť zasahuje seba samú)
PRODUKTOR	PDK	(produkuje zmyslovo vnímateľné produkty)
ELEMENTÍV	ELE	(nevedomá aktívna prírodná sila, živeľ)
STATUÁL	STA	(nositeľ stavu)
PACIENS	PAC	(zasahovaná entita)
REZULTANT	REZ	(vzniknutá entita)

DEŠTRUKTANT	DTK	(zaniknutá entita)
BENEFAKTOR	BEN	(v [ne]prospech koho)
MATERIÁL	MAT	(látka)
INŠTRUMENT	INS	(nástroj)
REALIZÁTOR	REA	(realizuje dej svojou vlastnou substanciou)
Verbálne procesy		
EDITOR	EDI	(pôvodca verbálneho procesu)
ADRESÁT	ADS	(adresát, komu)
FORMATÍV	FRM (	forma, čo)
TEMATÍV	TEM	(téma, o čom)
Majetkoprávne procesy		
DONOR	DON	(darca)
RECIPIENT	REC	(prijímateľ)
OBJEKT	OBJ	(predmet transakcie)
ANTIOBJEKT	ABJ	(protihodnota)
Percepčné a kognitívne procesy		
KOGNIZANT	KOG	(poznávateľ)
FENOMENÁL	FEN	(poznávané)
Citové procesy		
AFEKTOR	AFE	(entita prežívajúca citový stav)
EVOKÁTOR	EVO	(entita vyvolávajúca citový stav)
Atributívne procesy		
ATRIBUTANT	ANT	(nositeľ vlastnosti)
ATRIBUTÍV	ATR	(vlastnosť)
KVALIFIKÁTOR	KVF	(hodnotiaci, pripisovateľ vlastnosti)
Ekvatívne procesy		
IDENTIFIKANT	IFT	(kto je stotožňovaný)
IDENTIFIKÁTOR	IFR	(s čím je stotožňovaný)
Miestne a časové pády		
LOKATÍV	LOK	(kde)
DESTINATÍV	DES	(kam)
ORIGINATÍV	ORI	(odkiaľ)
VIÁL	VIA	(kadiaľ)
TERMINÁL	TER	(pokiaľ)
EXLOKATÍV	EXL	(odkadiaľ)

DIFERENCIÁL	DIF	(o koľko)
TEMPORÁL	TMP	(kedy)
INICIÁL	INC	(odkedy)
FINÁL	FIN	(dokedy)
FREKVENTÁL	FRE	(ako často)
DURATÍV	DUR	(ako dlho)
PERSPEKTÍV	PER	(nakedy)
Doplnkové pády		
KAUZATÍV	KAZ	(príčina)
KONZEKVENT	KZK	(následok)
MOTIVANT	MOT	(dôvod)
KONDICIONÁL	KND	(podmienka)
KONCESÍV	KCS	(prípustka)
RESPEKTÍV	RES	(zreteľ)
RELÁTOR	REL	(objekt prirovnania)
SOCIATÍV	SOC	(sprievodná okolnosť)
SUBSTITUÁL	SUB	(zastúpenie, za koho)
SUKCESÍV	SUK	(pôvod, po kom)
POSESÍV	POS	(vlastnícky vzťah)
MESURATÍV	MES	(miera)
MODUS	MDS	(spôsob)
VERIFIKATÍV	VER	(istotná modálnosť)
PARTITÍV	PAR	(časť celku)
OKUPATÍV	OKU	(objekt záujmu)
KOHERENTÍV	KOH	(neurčitá súvislosť)
Číslovkové pády		
KARDINÁL	KAR	(koľko)
ORDINÁL	ORD	(koľký)
KOLEKTÍV	KOL	(koľkoro)
MULTIPLIKATÍV	MUL	(koľkonásobný)
SPECIATÍV	SPE	(koľkoraký)
DISTRIBUTÍV	DIS	(po koľko)
FRAKTÁL	FRA	(koľká časť)

5.4 K pôvodu sémantických pádov

Pre zavedenie súčasného chápania pojmu sémantického pádu vo výpočtovej lingvistike bola rozhodujúca Fillmorova práca (1968). Horeuvedený pádový systém

sa vykryštalizoval na pozadí skúsenosti so systémom SAPFO. Pritom sme čiastočne zohľadnili práce iných autorov a pripomienky J. Horeckého. Niektoré pády, ako BENEFAKTOR, pochádzajú priamo od C. J. Fillmora (ibid.). Voľba osobitného systému pádov pre verbálne, kognitívne, atributívne a ekvatívne procesy bola ovplyvnená angloamerickou literatúrou (Winograd, 1983). Miestne a časové pády sme zvolili tak, aby prirodzeným spôsobom zodpovedali slovenským opytovacím zámenám. Pády PROCESOR, PRODUKTOR, nositeľ stavu a REALIZÁTOR sme prevzali od E. Tibenskej (1988).

Náš pádový systém sa do veľkej miery kryje so systémom 31 sémantických pádov, ktoré navrhli P. Sgall, E. Hajičová a J. Panevová (1986), iba niektoré pády chápeme odlišne alebo ich jemnejšie rozčleňujeme.

5.5 Problém voľby sémantických pádov

Voľba sémantických pádov je zdôvodnená a ospravedlnená princípom kontrastívnej diferenciácie (0.16). Napriek tomu obsahuje istý subjektívny prvok. Rozliční autori navrhujú dosť odlišné pádové systémy.

Pre nás však ďalej nebude také dôležité to, *ktorý* konkrétny pádový systém zvolíme, ako to, že musíme zvoliť *nejaký*, aby sme mohli testovať jeho správanie pri parafrázovaní a inferenciách. Okrem toho rátame s tým, že skutočné aplikácie dialógových systémov sa budú dať vybudovať iba nad užšími svetmi s konečno-doménovou sémantikou (Palmer, 1990), v ktorých je voľba pádového systému už viac alebo menej zrejmá. Napríklad v systéme GUS pre prácu s letenkami (Bobrow et al., 1977) to boli pády ako LIETADLO, LETISKO, PRÍLET, ODLET, FAJČIAR, OKNO a pod.

5.6 Sekundárna špecifikácia sémantických pádov

Výrazy *nad stolom, pod stolom, okolo stola* všetky spadajú pod LOKATÍV, avšak neradno ich zamieňať. Z tohto dôvodu pripájame k miestnym, časovým a niektorým ďalším pádom ešte sekundárnu špecifikáciu, ktorú budeme zapisovať malým číselným indexom. Potom *pod stolom* bude LOK_3, *nad stolom* LOK_2, *okolo stola* LOK_{13}. Pádový systém sa tým zjemní na 88 pádov.

Jemnejšia špecifikácia sémantických pádov		
Priestorové:	**Časové:**	**Emotívne zafarbenie:**
1. v, vnútri	**1.** pred	**1.** kladné
2. nad	**2.** tesne pred	**2.** záporné
3. pod	**3.** začiatkom	

4. pred	**4.** počas	**Sprievodná okolnosť:**	
5. za	**5.** koncom		
6. okolo	**6.** tesne po	**1.** s okolnosťou	
7. pri, vedľa	**7.** po	**2.** bez okolnosti	
8. na (povrchu)	**8.** práve keď		
9. vľavo od		**Miera:**	
10. vpravo od	**Podmienka**:		
11. medzi		**1.** malá	**4.** primalá
12. oproti	**1.** reálna	**2.** stredná	**5.** optimálna
13. mimo, vonku	**2.** ireálna	**3.** veľká	**6.** priveľká

5.7 Príklady

ELE DTK PCS PCS DES1
Vietor rozbil okno. Janko s Marienkou idú do lesa.

PDK STA AGS PAC INS
Ruže voňajú. Hrad spustol. Robotník nahadzuje uhlie lopatou.

AGS REZ MAT DTK REZ
Otec vystrúhal figúrku z dreva. Pohár sa rozbil na črepiny.

AGS PAC REA KOG FEN
Družstevníci naplnili voz obilím. Orfeus sa započúval do hudby.

AGS BEN1 PAC
Janko daroval Marienke kvety.

DON REC OBJ ABJ
Peter predal Pavlovi dom za mešec dukátov.

EDI ADS FRM TEM
Víla rozpovedala deťom rozprávku o princeznách.

KOG FEN AFE1 EVO
Buddha spoznal sám seba. Janko miluje Marienku.

KVA ANT ATR
Pápež vyhlásil Františka z Assisi za svätého.

IFT IFR LOK1
Dr. Cimrman sa stal riaditeľom závodu 01. Všetci sú na zámku.

ORI1 VIA7 DES7
Vyjdite z lesa a choďte popri potoku až k chalúpke.

DIF EXL TER
Posunul sa o dva metre. Od Žiliny po Trenčín sa spievalo.

INC FIN FRE TMP4
Od pondelka do piatku pracoval. Denne prší. Prídem zajtra.

MOT KND REL
Idem po chlieb. Za rosy sa dobre kosí. Je sladká ako jahoda.

KCS KAZ KZK
Hoci sme sa pripravili, dážď spôsobil veľké škody.

RES AGS VER MOD FAZ
Vzhľadom na váš stav, by sme azda mali začať liečbu.

SUK SOC2 KOH
Opevnenie po Rimanoch. Dokázal to bez pomoci. Nobelova cena.

POS PAR EDI SUB
Otcov kabát. Strecha domu. Ivan hovoril namiesto Petra.

MES VER OKU
Na smrť zaľúbený. Možno príde. Zaoberá sa astronómiou.

KAR ORD KOL SPE
Tri zlaté vlasy. Jano bol prvý. Štvoro topánok. Storaké zázraky.

DIS FRA MUL
Po desať zlatých. Osminka masla. Dvojnásobný víťaz.

5.8 Sémantika slovesných fráz

Predpokladajme, že máme k dispozícii výsledok morfologickej a syntaktickej analýzy. Našou úlohou je sformulovať explicitné pravidlá, algoritmus, podľa ktorého bude počítač schopný mechanicky priraďovať sémantické roly jednotlivým aktantom. Aké gramatické, syntaktické a lexikálno-sémantické príznaky je nevyhnutné a postačujúce brať do úvahy, aby sme dokázali rozhodnúť o role participanta pri slovese?

Rola participanta závisí od štyroch hlavných faktorov:

a) Od intencie centrálneho slovesa:

LOK
stáť nad niečím (kde)

FEN
zamyslieť sa nad niečím (predmet poznávania)

b) Od predložky:

DES1 ORI1
ísť do lesa (kam) *ísť z lesa* (odkiaľ)

c) Od gramatického pádu substantíva:

SUB LOK
písal za Karola (namiesto) *písal za Karolom* (kde)

TMP4
písal za Karola IV. (kedy)

d) Od sématického príznaku participanta:

INS VIA
letieť lietadlom (čím) *letieť oknom* (kadiaľ)

MDS
letieť veľkou rýchlosťou (ako)

e) Od iných komplexných faktorov; syntaxe, slovosledu a intonácie:

POS Ivan nemá zbraň.	KOG Ivan nemá záujem.
LOK Nad lesom sa zamyslel.	FEN Zamyslel sa nad lesom.

Podľa toho pravidlá pre priraďovanie sémantických rôl participantom budú vyzerať nasledovne:

intenčný typ			sémantický príznak		sémantický pád
slovesa	+ predložka +	gramatický pád +	participanta	=	(rola)

Napríklad:

pohybové typu *ísť* + do + genitív + lokálium = destinatív$_1$

Funkciu, ktorá definuje vzťah medzi formou a významom v slovesných frázach budeme volať *valenčnou funkciou* (VAL) a zapisovať ju v tvare *sémantických rovníc* alebo predikátov:

VAL(1,do,gen,lok) = DES$_1$
VAL(1,do,gen,lok,DES$_1$).

5.9 Valencia

Pojem *valencia* je prevzatý z chémie (Tesniére, 1959). Chemické prvky majú v poslednej, tzv. *valenčnej* elektrónovej sfére voľné miesta pre určitý počet elektrónov. Valenčná sféra je rozhodujúca pre špecifické väzobné vlastnosti prvkov. Tiež každé sloveso má voľné miesta pre aktanty určitého druhu, čím je daná jeho špecifická spájateľnosť. Valenčný systém slovesa pozostáva z troch sfér (porov. aj Daneš, 1971):

1 sféra:	obligátne doplnenia	intencia slovesa	valencia slovesa
2 sféra:	špecifické doplnenia	fakultatívne doplnenia	
3 sféra:	všeobecne pripojiteľné doplnenia		

5.10 Intenčné rámce

Súpis všetkých prípustných doplnení, ktoré sú pre sloveso špecifické, sa nazýva *intenčným rámcom* slovesa. Napríklad intenčný rámec slovesa *povedať* vyzerá takto:

0	+	nom	=	EDI (živ.)
0	+	ak	=	FRM
0	+	dat	=	ADS
o	+	lok	=	TEM
za	+	ak	=	SUB (živ.)
0	+	inš	=	INS (inštr.), MDS (abstr.)
s	+	inš	=	EDI (živ.), MDS (abstr.)
cez	+	ak	=	INS (inštr.), VIA_1 (lok.)
proti	+	dat	=	RES_2
z	+	gen	=	ORI_1
do	+	gen	=	DES_1
bez	+	gen	=	SOC_2
že, či	+	predm.v.	=	TEM
aby	+	predm.v.	=	MOT

Sémantické príznaky substantív sú uvedené v zátvorkách vpravo ako podmienky, ktoré musí aktant splniť, aby mohol stáť v príslušnej role. Nad prvou čiarou sú uvedené povinné doplnenia, pod čiarou fakultatívne.

Pod druhou čiarou sú uvedené nenominálne, ale intenčne špecifické valenčné miesta. Napríklad slovesá typu *vidieť* priberajú aj infinitív a vedľajšiu vetu predmetovú. Aj nenominálny aktant môže patriť medzi obligátne. Napríklad príslovka pri slovese *tvárit' sa (ako)*.

Pri tvorbe intenčných rámcov stačí zaoberať sa iba holými gramatickými pádmi a predložkovými pádmi prvotných predložiek (*za, o, po, na, v, nad, pod, pred, medzi, s, cez, pre, k, proti, bez, od, do, z, u*). Druhotné predložky majú svoj špecifický význam samy osebe, ktorý je invariantný vzhľadom na slovesá (5.14). Len význam prvotných predložiek sa vyhraňuje až v kontexte.

Prvotné predložkové pády sa ďalej delia na tie, ktoré sú vzhľadom na slovesá špecifické, a tie, ktoré sú vzhľadom na ne invariantné. Napríklad *v + lokál* je príslovkové určenie miesta, ktoré možno pridať ku každému slovesu bez výnimky. Takéto pády taktiež nepatria do intenčného systému slovesa a možno ich vypísať osobitne.

Predložkové pády, ktoré nie sú v intenčnom rámci uvedené (a nie sú ani spoločné všetkým slovesám), sloveso nevytvára.

5.11 Triedy ekvivalencie slovies

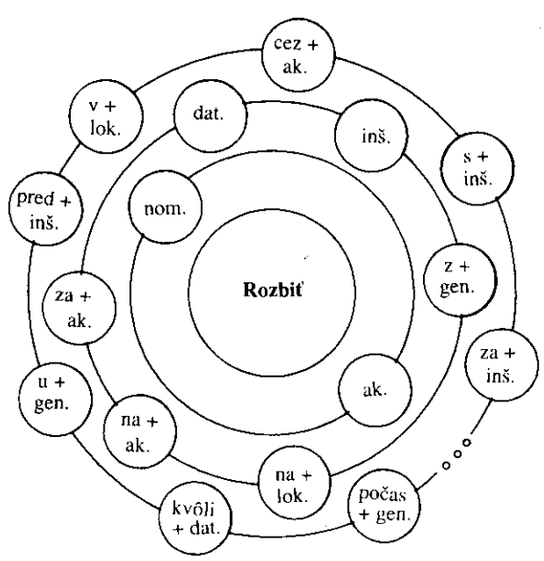
Vypíšme všetky intenčné rámce slovenských slovies. Zistíme, že rámcov bude menej ako slovies, pretože slovesám s príbuznou sémantickou štruktúrou budú prislúchať rovnaké intenčné rámce. Napríklad do intenčného rámca slovesa *ísť* padnú aj slovesá ako *bežať, cestovať, letieť, pochodovať, liezť, plaviť sa, posúvať sa, cválať*... Stačí teda vypísať len jeden rámec pre triedu všetkých intranzitívnych vedomých translačných slovies.

VALENČNÝ SYSTÉM SLOVESA

Kvôli názornosti si štruktúru slovesa môžeme predstaviť ako chemický prvok: každé sloveso má tri „valenčné sféry, na ktorých má voľné miesta pre doplnenia určitého typu. Všetky tri sféry spolu tvoria ***valenčný systém****, t. j. všetky možné doplnenia, aké môže sloveso prijímať. Z toho vnútorné dve sféry tvoria* ***intenčný systém*** *- doplnenia, ktoré sú pre sloveso typické. Na najvrchnejšej sfére sú iba tie doplnenia, ktoré môžu prijímať viac-menej všetky slovesá, t. j. nie sú pre intenciu daného slovesa špecifické.*

Intenčný systém sa ešte delí na doplnenia obligátne (na najvnútornejšej sfére) a fakultatívne (na druhej sfére zvnútra). Zatiaľ čo neprítomnosť povinného pádu sa pri slovese pociťuje ako chybná formulácia, z fakultatívnych pádov si môžeme (a nemusíme) vyberať. K obligátnym pádom nemusia patriť len hlavné vetné členy, ale aj iné, napr. MODUS pri slovese správať sa ***ako****.*

Obr. 7



Na množine slovenských slovies zavedieme takúto binárnu reláciu: *Dve slovesá sú ekvivalentné práve vtedy, keď majú rovnaký intenčný rámec.* Ľahko nahliadneme, že vzťah je reflexívny, symetrický a tranzitívny, tvorí teda ekvivalenciu a definuje rozklad množiny slovies na disjunktné triedy.

Uvedenou metódou sme systematizovali slovesá z Krátkeho slovníka slovenského jazyka (1987). Je ich tam okolo 9000. Dostali sme 120 slovesných tried ekvivalencie (Páleš, 1988). Množstvo slovies sa vymyká systematizácii svojimi ojedinelými vlastnosťami. Patria medzi ne napríklad slovesá s predložkovými pádmi motivovanými vizuálne (*sliediť za dakým, vládnuť nad niekým...*) alebo archaické väzby. V najhoršom prípade z nich možno utvoriť triedy s malým počtom slovies.

5.12 Niektoré slovesné intenčné rámce

20. Manuálny tranzitívny proces konštruktívny

0	+	nom	=	AGS (živ.), INS (inštr.)
0	+	ak	=	REZ
z,do	+	gen	=	MAT
0	+	dat	=	BEN
0	+	inš	=	INS (inštr.), MDS (abstr.)
na	+	lok	=	INS (inštr.)
pre	+	ak	=	BEN
na	+	ak	=	MOT
bez	+	gen	=	SOC
za	+	ak	=	SUB (živ.)
za	+	gen	=	KND (prír.)
proti	+	dat	=	MOT
s	+	inš	=	AGS (živ.), INS (inštr.)
pri	+	lok	=	KND (prír.)
čo	+	predm.	=	REZ

vyrobiť, produkovať, vystrúhať, aranžovať, (u)variť, vyrezávať, (vy)brúsiť, odliať, vysekať, vypíliť, raziť (mince)...

1.Translačné intranzitívne vedomé procesy

0	+	nom	=	PRO
do	+	gen	=	DES_1 (lok.)
k	+	dat	=	DES_7 (lok.)
0	+	inš	=	INS (inštr.), VIA1 (lok.), MDS (abstr.)
na	+	ak	=	DES_8 (lok.), MOT (abstr.)

nad	+	ak	=	DES_2 (lok.)
pod	+	ak	=	DES_3 (lok.)
cez	+	ak	=	VIA_1 (lok.)
po	+	ak	=	MOT (živ.; inštr.), TER_7 (lok.)
pre	+	ak	=	MOT
od,z	+	gen	=	$ORI_{7,1}$ (lok.), KAZ (abstr.)
za	+	inš	=	DES (živ.), VIA_5 (mov.)
za	+	ak	=	DES_5 (lok.), SUB (živ.)
0	+	dat	=	POS
0	+	ak	=	MES (jedn.)
v	+	lok	=	INS (inštr.), SOC (odev), KND (prír.)
pri	+	lok	=	VIA_7 (lok.)
nad	+	inš	=	VIA_2 (lok.)
pod	+	inš	=	VIA_3 (lok.)
po	+	lok	=	VIA_8 (lok.)
za	+	gen	=	KND (prír.), TMP (živ.)
o	+	ak	=	DIF (jedn.)
s	+	inš	=	PRO (živ.), INS (inštr.), SOC (abstr.)
proti	+	dat	=	DES_{12} (lok.)

ísť, bežať, cestovať, letieť, pochodovať, liezť, plaviť sa, posúvať sa, cválať...

32. Reflexívny materiálny proces deštruktívny

0	+	nom	=	DTK
na	+	ak	=	REZ
od	+	gen	=	KAZ (prír.)
0	+	dat	=	BEN, POS (živ.), PAR (neživ.)
za	+	gen	=	KND (prír.)
pri	+	lok	=	KND (prír.)
pod	+	inš	=	ELE (prír.)
0	+	inš	=	KAZ (prír.), MDS (abstr.)

zhoriet', rozpadnúť sa, rozdrobiť sa, (roz)topiť sa, rozplynúť sa...

80. Kognitívny proces s akuzatívom

0	+	nom	=	KOG (živ.)
0	+	ak	=	FEN
0	+	inš	=	MDS (abstr.), INS
0	+	dat	=	BEN
u	+	gen	=	POS (živ.)
za	+	ak	=	SUB (živ.)

že,čo	+	predm.	=	FEN
kto	+	podm.	=	KOG (živ.)

analyzovať, predstaviť si, uvažovať (koho čo), uvážiť, poznávať, určiť, tušiť, zamýšľať, mýliť si, uvedomiť si, anticipovať, akceptovať, skúmať, zaumieniť si...

50. Percepčný proces orientovaný

0	+	nom	=	PER
0	+	ak	=	FEN
na	+	ak	=	DES_8 (lok.)
do	+	gen	=	DES_1 (lok.)
nad	+	ak	=	DES_2 (lok.)
pod	+	ak	=	DES_3 (lok.)
pred	+	ak	=	DES_4 (lok.)
0	+	dat	=	POS (živ.), PAR (neživ.)
cez	+	ak	=	VIA_1 (lok.), INS (inštr.)
od,z	+	gen	=	$ORI_{7,1}$ (lok.)
0	+	inš	=	INS (inštr.), VIA_1 (lok.)
k	+	dat	=	DES_7 (lok.)
bez	+	gen	=	SOC_2
u	+	gen	=	POS (živ.)
za	+	ak	=	DES_5 (lok.)
po	+	ak	=	TER_7 (lok.)
do	+	gen	=	DES_1 (lok.)
s	+	inš	=	PER (živ.), INS (inštr.)
0	+	inf.	=	FEN
že,či	+	predm.	=	FEN
kto,čo	+	predm.	=	FEN
kto	+	podm.	=	PER (živ.)

vidieť, počuť...

84. Mentálny proces s „v“

0	+	nom	=	KOG (živ.)
v	+	ak	=	FEN
na	+	ak	=	FEN
že	+	predm.	=	FEN

veriť, dúfať...

105. Citový proces s inštrumentálom

0	+	nom	=	AFE (živ.)
0	+	inš	=	EVO_2
pre	+	ak	=	EVO_2

trápiť sa, zožierať sa...

106. Citový proces s „do"

0	+	nom	=	AFE (živ.)
do	+	gen	=	EVO_1
pre	+	ak	=	KAZ
0	+	inš	=	MDS (abstr.)

zaľúbiť sa...

140. Zmena stavu

0	+	nom	=	STA
od,z	+	gen	=	KAZ
po	+	lok	=	KAZ
0	+	dat	=	BEN, POS (živ.), PAR (neživ., korp.)
za	+	gen	=	KND
0	+	inš	=	MDS (inštr.)

(s)červenieť, (z)ľadovatieť, drevnatieť, (s)krásnieť, chorľavieť, glejovatieť, deravieť, jasnieť, pustnúť, uzdraviť sa, (z)vlhnúť, (vy)triezvieť, (s)chudobnieť, (s)kamenieť, zhavranieť...

180. Atmosférické prírodné javy orientované

do	+	gen	=	DES_1 (lok.)
na,za	+	ak	=	$DES_{8,5}$ (lok.)
nad,pod	+	ak	=	$DES_{2,3}$ (lok.)
cez	+	ak	=	VIA_1 (lok.)
0	+	dat	=	BEN, POS (živ.), PAR (neživ., korp.)
0	+	inš	=	MDS (rozvité subst.)
k	+	dat	=	DES_7 (lok.)
z	+	gen	=	ORI_1 (lok.)
po	+	ak	=	TER_7 (lok.)

sneží, leje, chumelí sa, udrelo, kadilo (sa), prášilo (sa), prší, mrholí, fúka, duje...

201. Atributívne procesy

0	+	nom	=	ANT	
0	+	dat	=	KVF	(živ.)
	+	adj	=	ABT	
aký	+	prívl. v.	=	ABT	

byť, zdať sa, pripadať, javiť sa, pozdávať sa, vidieť sa, (z)ostať...

205. Ekvatívny proces s „na"

0	+	nom	=	KVF	(živ.)
0	+	ak	=	IFT	
z	+	gen	=	IFR_1	
na	+	ak	=	IFR_2	
za	+	ak	=	KAZ	

povýšiť, degradovať...

400. Verba essendi

0	+	nom	=	AGS	(živ.)

kraľovať, chalupárčiť,

5.13 Pádové preferencie

Nepovinné predložkové pády slovesa sa líšia v miere svojej „očakávanosti", „predpokladanosti". Niektoré sú priam typické, iné sú možné, ale vyskytujú sa len zriedka. Napríklad pri slovese *ísť* nás ako prvé napadne *kam*, pri *odísť - odkiaľ*, *rozprávať - o čom*. *Ísť medzi koho čo, rozprávať cez koho/čo* nie sú frekventované.

Aby sme túto informáciu o frekvencii predložkových pádov mohli využiť, zoradíme riadky v intenčných rámcoch tak, že frekventované pády budú na začiatku a menej frekventované na konci. Neskôr (v kapitole 15) zavedieme ku každému pádu osobitné číselné ohodnotenie, ktoré bude vyjadrovať jeho frekvenčný podiel na valencii slovesa. To nám umožní zadefinovať tzv. *preferovaný pád*. Preferovaný pád je najfrekventovanejší sémantický pád, ktorý unifikuje s danou sémantickou črtou.

Pojem preferovaných pádov nám umožní rekonštruovať sémantickú štruktúru fráz, v ktorých sa časť informácie stratila kondenzáciou. Po nominalizácii alebo

kondenzácii slovesnej syntagmy do kompozita či derivátu vzťah medzi slovesom a aktantom už nemusí byť vyjadrený explicitne:

AGS PAC AGS PAC
zločincov útek - zločincovo zatknutie, ľudovláda - svetovláda

INS LOK TEM REZ
delostrelec - horolezec - zemepisec - zvonolejec

No súčinnosťou derivatologického modulu, kolokačných funkcií a frekvenčno-preferenčnej informácie sa darí vzťah rekonštruovať (pozri príklady 5.25, 5.26, 7.10 až 7.15).

5.14 Sémantika druhotných predložiek. Nešpecifické doplnenia

Druhotné predložkové pády a nešpecifické doplnenia sa dajú pripájať ku všetkým druhom slovies a ich význam je od slovesa nezávislý. Nie sú teda súčasťou intenčného rámca slovesa, a stačí ich uviesť osobitne (Páleš, 1988). Uvedieme len niektoré:

v + lok =	LOK_1 (lok.)	po + lok =	TMP_7 (čas.)
u + gen =	LOK_7 (lok.)	za + gen =	TMP_4 (živ.)
na + lok =	LOK_8 (lok.)	pred + inš =	TMP_1 (čas.)
za + inš =	LOK_5 (lok.)	cez + ak =	TMP_4 (čas.)
pri + lok =	LOK_7 (lok.)	bez + gen =	SOC_2

ponad, popod, popred, poza, pomedzi + ak,inš = VIA_i, DES_i, LOK_i
znad, spred, spoza, spopod, zôkol + gen = $ORI_{2,4,5,3,6}$

vôkol + gen =	LOK_6 (lok.)	vďaka + dat =	KAZ
naprieč + inš =	VIA_1 (lok.)	následkom + gen =	KAZ
napriek + dat =	KCS	namiesto + gen =	SUB
kvôli + dat =	MOT	v súlade s + inš =	RES_1
za účelom + gen =	MOT	s ohľadom na + ak =	RES_2
na úkor + gen =	BEN_2	pomocou + inš =	INS (ins.)

5.15 Počítadlo povinnej intencie

SAPFO používa ešte jedno kritérium správnej formulovanosti viet. Obligátne aktanty pri slovese nesmú chýbať ani byť navyše.

Napríklad pri tranzitívnych slovesách je povinný práve jeden predmet v akuzatíve. Inak sa veta pociťuje ako zle formulovaná. Aby sme nemuseli vypisovať všetky prípustné kombinácie, priradíme každému predložkovému pádu číslo z intervalu <0,1>, ktoré vyjadruje jeho *podiel na povinnej valencii* centrálneho slovesa:

ťahať

0	+	ak	=	OBJ	(konkr.)	1
0	+	inš	=	INS	(inštr.)	0
do	+	gen	=	DES	(lok.)	0

začleniť

0	+	ak	...	3/4
do	+	gen	...	1/4
k	+	dat	...	1/4
medzi	+	ak	...	1/4
ostatné pády			...	0

Čísla sú volené tak, aby suma „podielov na valencii" bola rovná jednej práve pri každom správnom výbere predložkových pádov.

Počítadlo postupne sčituje valenčné podiely jednotlivých aktantov, ktoré vo vete stretá. Na konci musí byť suma S = 1. Ak S = 1, znamená to, že povinná valencia ešte nebola naplnená (S<1) alebo bola prekročená (S>1). Počítadlo vyhlási vetu za neakceptovateľnú.

5.16 Sémantika prísloviek

Každá príslovka má svoju určitú sémantickú rolu, ktorá vyplýva z jej lexikálneho významu a ktorú nemení. Táto rola je daná slovníkovo, priamo ako sémantický príznak prísloviek.

MDS	dôstojne, krásne, obratne, presne, veselo, pomaly
RES	telesne, spoločensky, finančne, mravne
MES	veľmi, veľa, málo, trochu, čiastočne, nadobro
LOK	doma, vonku, vzadu, vysoko, severne
DES	domov, nahor, dopredu
ORI	odspodu, zdola, zdnuka
VIA	naprieč, zadom, krížom-krážom
TMP	dnes, zajtra, večer, popoludní
PSP	nazajtra, nateraz
DUR	chvíľu, stále, ustavične, večne, nakrátko
FRE	denne, občas, pravidelne
INC	oddnes, odjakživa
FIN	doteraz, dozajtra, doživotne
KAZ	náhodou, omylom, úmyselne
MOT	bezcieľne, pracovne

5.17 Sémantika vedľajších viet

Ako aktant podradený slovesu môže vystupovať aj celá vedľajšia veta. V takom prípade sémantická rola závisí od typu vedľajšej vety a od vzťažného zámena, ktorým je uvedená.

Predmetové, podmetové, vetnozákladové a prísudkové vedľajšie vety nachádzame len pri slovesách istých typov (dicendi, sentiendi, percipiendi a sponové slovesá). Sú uvedené aj so svojimi sémantickými rolami v intenčných rámcoch príslušných slovies (5.10, pod druhou čiarou).

Roly *príslovkových* vedľajších viet nezávisia od slovies; ich sémantickú rolu identifikujeme len podľa zámena alebo spojky, ktorá ich uvádza.

Typy príslovkových vedľajších viet a ich sémantické roly

Sémantická rola	*Uvádzacie zámená a spojky*	*Odkaz v hlavnej vete (nepovinný)*
LOK	kde, kdekoľvek	tam, všade, ta
VIA	kade, kadiaľ	tadiaľ, tade, všade
ORI	skade, odkiaľ	stade, odtiaľ
DES	kam	tam, všade
TER	pokiaľ, pokadiaľ	potadiaľ, potiaľ
EXL	odkadiaľ	odtadiaľ
TMP_4	keď, ako, kým, zatiaľ čo	vtedy
FRE	kedykoľvek, čo raz, čo	vždy
TMP_1	skôr ako, prv ako, kým	
TMP_6	ako, len čo, iba, sotva	už
INC	ako, čo	odvtedy, od tých čias
FIN	kým, až, dokiaľ	dotiaľ, dovtedy
KAZ	lebo, keď(že), pretože	preto, zato
MOT	aby, (že)by, nech,	nato, zato
KND_1	ak(že), keď, pokiaľ	tak, nuž, potom, teda
KND_2	keby, by, ak by	
KCS	hoci, čo, hoc aj, čo ako	predsa, jednako
MDS	ako(by), sťa	
REL	ako, sťa by	tak
MES	ako, koľko, čím	toľko, čím
KZK	takže, až, že, aby tak	toľko, taký

5.18 Kontrastívny príklad

Analyzujme syntakticky rovnorodé frázy:

letieť lietadlom *letieť oknom* *letieť veľkou rýchlosťou*

Ak slovník obsahuje (časť slovotvorných hniezd a dištinktívne sémantické príznaky):

letieť	(1)	okno	(lok)	rýchly	(kval)
let	(dej)	okenný	(klas)	rýchlo	(kval)
lietadlo	(inštr)	okenica	(lok)	rýchlosť	(atr)
letuška	(osoba)	oknár	(osoba)	rýchlik	(stroj)

Aktivujú sa tieto tri riadky z intenčného rámca [1] pre vedomé tranzitívne translačné procesy:

VAL(1,0,ins,instr,INS).
VAL(1,0,ins,lok,VIA_1).
VAL(1,0,ins,abs,ATR).

Dostaneme túto výslednú analýzu:

let-
INS» lietadl-

let-
VIA_1» okn-

let-
ATR» rýchl- $^{+mag}$

V tomto prípade o správnych sémantických rolách rozhodli dištinktívne sémantické príznaky participantov. *Lietadlo* je NÁSTROJ, *okno* PRIESTOR a *rýchlosť* SPÔSOB pohybu.

5.19 Kontrastívny príklad

Všimnite si rozdiel v reprezentáciách nasledujúcich dvoch syntakticky rovnorodých viet. V prvom prípade je vedľajšia veta participantom hlavnej vety vyjadrujúcim čas (TEMPORÁL). V druhom prípade vedľajšia veta vyjadruje tiež čas, ktorý sám vystupuje vzhľadom na hlavnú vetu v role objektu poznávania (FENOMENÁLu). Tým vzniká v sieti nepomenovaný, neznámy vrchol, ktorý je definovaný iba tým, že sa naň obe vety vzťahujú a on ich obe spája. (V prvom prípade ide o vedľajšiu vetu *príslovkovú*, v druhom o *predmetovú*.)

Ivan príde, kedy sa mu zachce.

ísť $^{+ad\ +fut}$
PCS» Ivan
TMP_4» chcieť $^{+fut}$
KOG» = PCS

Peter nevie, kedy Ivan príde.

vedieť $^{+neg}$
KOG» Peter
FEN» **X**
«TMP ísť $^{+ad\ +fut}$
PCS» Ivan

5.20 Protipríklad

V niektorých prípadoch je na pravej strane intenčných pravidiel uvedených viac sémantických pádov, a ani dištinktívne príznaky substantív neposkytujú dostatočnú

informáciu pre jednoznačné rozhodnutie. Napríklad *choď za Ivanom* môže byť DESTINATÍV (k nemu domov) ale aj VIÁL (v jeho pätách). *Utekaj po kolík* môže znamenať MOTIVANT (účel) ale aj TERMINÁL (pokiaľ).

V takom prípade odovzdáva funkcia VAL analýzu na ďalšie spracovanie ako viacznačnú.

5.21 Sémantika nominálnych fráz

Nominálne frázy vystupujú ako aktanty okolo slovesa v sémantickej štruktúre vety a zároveň sú samy štruktúrované obdobným spôsobom. Menná fráza sa skladá z hlavného substantíva, ku ktorému sa zľava a sprava pripájajú zhodné a nezhodné prívlastky. Každý prívlastok vystupuje vzhľadom k hlavnému substantívu v určitej sémantickej role (páde).

Tieto roly nestačí kvalifikovať len ako 'atribúty'. Napríklad reprezentácia genitívnych prívlastkov

fajčenie cigariet	*fajčenie žien*
fajč-	fajč-
ATR» cigaret-	ATR» žen-

len pomocou pádu ATRIBÚT by nebola dostatočná. Až keď nahliadneme, že v prvom prípade je prívlastok vzhľadom na hlavné meno v role PACIENSa a v druhom v role AGENSa, budeme schopní tieto menné frázy parafrázovať aj slovesne. Inak by sme tieto dva rozdielne prípady zamieňali:

fajč-	fajč-
PAC» cigaret-	AGS» žen-
cigarety sú fajčené (ženami)	ženy fajčia (cigarety)

5.22 Nominalizácia

Podobne ako slovesá aj substantíva majú svoju špecifickú zamerenosť a spájateľnosť (1.3) a vytvárajú bohatý valenčný systém rôl. Prívlastky majú rovnako veľkú škálu sémantických pádov ako slovesné aktanty.

Deverbatíva preberajú valenčný systém slovies, od ktorých sú odvodené. Nominálne frázy s deverbatívnym hlavným menom možno chápať ako nominalizované vety. Jednotlivé aktanty sa premenia na zhodné a nezhodné prívlastky, pričom po nominalizácii zostáva systém sémantických rôl nezmenený:

PCS TMP MDS DES
Sused včera rýchlo cestoval do Prahy →

PCS TMP MDS DES
susedova včerajšia rýchla cesta do Prahy.

Neodvodené, desubstantívne a deadjektívne substantíva majú tiež svoje valenčné vlastnosti, ktoré sú dané sémantickou črtou substantíva a jeho koreňa. Aj tieto menné frázy môžeme chápať ako výsledok nominalizácie, ak predpokladáme implicitnú prítomnosť slovesa *byť*:

ATR ORI TMP
stará Kremnická zlatka zo 14. storočia

REL PAR MAT SUK
špicaté cimburie hradieb *kamenné opevnenie po Rimanoch*

5.23 Funkcie LVAL, RVAL

Od čoho závisí sémantická rola prívlastku pri substantíve? Pri zhodných prívlastkoch sme identifikovali dvanásť relevantných argumentov:

1) Prívlastok (adjektiválium).
2) Sémantická črta prívlastku.
3) Derivačný prefix prívlastku.
4) Derivačný sufix prívlastku.

5) Slovný druh koreňa, od ktorého je prívlastok odvodený.
6) Sémantická črta koreňa prívlastku.

7) Hlavné substantívum.
8) Jeho sémantická črta.
9) Jeho derivačný prefix.
10) Jeho derivačný sufix.

11) Slovný druh koreňa, od ktorého je hlavný člen odvodený.
12) Sémantická črta koreňa.

Nezhodné prívlastky, ktoré sa nepripájajú zhodou zľava ale primkýnaním sprava, majú relevantných argumentov desať:

1) Hlavné substantívum.
2) Jeho sémantická črta.

3) Slovný druh koreňa, od ktorého je hlavný člen odvodený.
4) Sémantická črta koreňa.

5) Gramatický pád prívlastku.
6) Predložka (spojka, vzť. zámeno), ktorou sa prívlastok pripája.

7) Slovný druh prívlastku.
8) Jeho sémantická črta.

9) Slovný druh koreňa, od ktorého je prívlastok odvodený.
10) Sémantická črta koreňa prívlastku.

12-árna kolokačná funkcia LVAL definuje ľavostrannú spájateľnosť substantív. LVAL o každom zhodnom prívlastku rozhodne, či môže pri substantíve stáť, a ak áno, priradí mu jeho sémantickú rolu.

LVAL je stelesnená v 33-och komplexných pravidlách (pozri programové moduly VIII, IX). Napríklad pravidlo 17 vyzerá takto:

```
LVAL((adj,akvzt,Pr,Sf),(sub,Se1),(sub,Se2,_,_),(Sd,_),SP) :-
    (Sf=n;Sf=sk;Sf=ov;Sf=ick;Sf=jš),
    ((Se1==lok,ifthenelse(sema(Se2,mov),SP=ori_i,Sp=lok_i);
      (Se1==cas,SP=tmp_i)),
    (Sd==nil;Sd==sub),
    index(Pr,Se1,i).
```

Toto pravidlo spracúva všetky adjektíva, akostno-vzťahové, odvodené slovotvorným prefixom *Pr* a jedným zo sufixov *-ný, -ský, -ový, -ický* alebo *-jší*, od substantíva, so sémantickou črtou *lok* alebo *čas*, pričom hlavný člen je substantívum so sémantickou črtou Se2, neodvodený alebo desubstantívny. Ak prívlastok je odvodený od miestneho substantíva, potom ak hlavné substantívum je premiestňovateľné, ide o pád ORIGINATÍV, ak je nepremiestňovateľné, ide o LOKATÍV. Ak je prívlastok odvodený od dejového substantíva, ide o TEMPORÁL. Malý index sémantického pádu sa určí podľa derivačného prefixu:

prímorská živica	*zadunajské pláne*	*predpotopný stroj*
živica	pláň $^{+pl}$	stroj
ORI_7» more	LOK_5» Dunaj	TMP_1» potopa

10-árna kolokačná funkcia RVAL definuje pravostrannú spájateľnosť substantív. RVAL o každom nezhodnom prívlastku rozhodne, či môže pri substantíve stáť, a ak áno, priradí mu jeho sémantickú rolu. RVAL je stelesnená v 36-tich komplexných pravidlách (pozri programové moduly VIII, IX). Napríklad pravidlo 4 vyzerá takto:

```
RVAL((sub,_),(sl,I),Pr,Pa,(sub,Se),_,SP) :-
    VAL(I,Pr,Pa,Se,SP).
```

Pravidlo prijíma všetky substantíva odvodené od slovesa s nejakou intenciou *I*, ku ktorým sa pomocou predložky *Pr* a gramatického pádu *Pa* pripája substantívum so sémantickou črtou *Se*. V takom prípade pravidlo zavolá 5-árnu funkciu VAL a odovzdá jej *I, Pr, Pa, Se* ako parametre. Funkcia VAL vráti sémantický pád, ktorý by tento prívlastok vytváral, keby stál ako aktant v intenčnom poli pôvodného slovesa. Pravidlo teda stelesňuje poznatok, že na deverbatívne nominálne frázy sa prenáša valenčný systém slovesa, od ktorého sú odvodené:

dar sestre

darovať
BEN» sestra

búchanie kladivom

búchať
INS» kladivo

cesta do Indie

cestovať
DES_1» India

rozprávka o Snehulienke

rozprávať
TEM» Snehulienka

5.24 Prívlastkové vedľajšie vety

Ak sa podstatné meno už nedá rozviť člensky, môže byť rozvité aj prívlastkovou vedľajšou vetou:

krajina, kde nezapadá slnko

krajina
«LOK_1 zapadať $^{+neg}$
AGS» slnko

kvet, ktorý kvitne iba raz

kvet
«AGS kvitnúť
MUL» jeden

predstava, že bude pršať

predstaviť si
FEN» pršať $^{+fut}$

informácia, prečo prišli Turci

informovať
TEM» **X**
«KAZ ísť $^{+ad\ +pret}$
PCS» Turek $^{+pl}$

Prívlastkové vedľajšie vety patria medzi nezhodné prívlastky a tiež sa spracúvajú v rámci formálnej funkcie RVAL. Napríklad prvú a tretiu frázu zanalyzovali pravidlá č. 31 a 32:

(31)
```
RVAL((sub,Se),_,0,Vzt,(sl,I),_,SP) :-
    ((Vzt=kde,SP=lok);
    (Vzt=kam,SP=des);
    (Vzt=odkiaľ,SP=ori)),
  VAL(I,0,Vzt,Se,SP),sema(Se,kon).
```

```
(32) RVAL((sub,dej),(sl,I),0,Spojk,(sl,_),_,SP) :-
        (Spojk=že;Spojk=či),
        VAL(I1,0,Spojk,_,SP).
```

5.25 Kontrastívny príklad

Analyzujme syntakticky rovnorodé frázy:

otcov odlet / otcovo zatknutie / otcova pozostalosť
otcova prísnosť / otcov nos/otcov kufor / otcova symfónia

Parafrázovať individuálne privlastňovacie prídavné mená *otcov, matkin* ako *patriaci otcovi, matke* je nedostatočné. To isté platí pre druhové privlastňovacie prídavné mená: *Labutí krk* nie je *krk patriaci labuti*, ale *krk ako labutí*; *orlie pero* nie je *pero patriace orlovi*, ale *pero z orla*; ani *lišací hon* nie je *hon patriaci líške* ale *hon na líšku*.

Fungovanie posesívnych adjektív definujú tieto ľavovalenčné nominalizačné rovnice:

```
LVAL((adj,indprivl,_,_),(sub,osoba),(sub,_,_,_),(sl,I),SP)
   :- prefer_pád(I,_,_,osoba,SP).

LVAL((adj,indprivl,_,_),(sub,osoba),(sub,atb,_,Sf),(adj,_),atr)
   :- (Sf=ota;Sf=oba;Sf=stvo;Sf=osť).

LVAL((adj,indprivl,_,_),(sub,osoba),(sub,Se,_,_),_,SP)
   :- ((Se=korp,SP=par);
       (Se=korp,sema(Se,kon),SP=pos);
       (SP=koh)).

prefer_pád(I,Pr,Pa,Se,SP) :- VAL(I,Pr,Pa,Se,SP),objekt_pád(SP),!.
prefer_pád(I,Pr,Pa,Se,SP) :- VAL(I,Pr,Pa,Se,SP),subjekt_pád(SP),!.
prefer_pád(I,Pr,Pa,Se,SP) :- VAL(I,Pr,Pa,Se,SP),!.
```

Nech slovotvorný a lexikálno-sémantický slovník (okrem iného) obsahujú:

zostať$_1$	(120)	letieť	(1)
zostatok	(kon)	priletieť	(10)
		odletieť	(10)
zostať$_2$	(301)	odlet	(dej)
pozostalosť	(kon)		

		zatknúť	(125)
cestovať	(1)		
$cesta_1$	(dej)	prísny	(kval)
cestovateľ	(osoba)	prísnosť	(atb)
$cesta_2$	(lok)	nos	(korp)
cestár	(osoba)		
cestný (akvzt)		kufor	(vec)
dediť	(300)	symfónia	(abs)
dedičstvo	(kon)		

Synonymický slovník (časť):

konverz(zostať$_2$,zdediť).
synon(cesta$_2$,vozovka).

A valenčný slovník obsahuje riadky:

VAL(10,0,nom,živ,ags).
VAL(10,z,gen,lok,ori).

VAL(301,0,nom,neživ,sta).
VAL(301,po,lok,živ,suk).

VAL(125,0,ak,živ,pac).
VAL(125,0,nom,živ,ags).
VAL(125,pre,ak,_,kaz).

VAL(1,0,nom,živ,ags).
VAL(1,do,gen,lok,des).

Potom použitím horeuvedených sémantických rovníc dostávame tieto parafrázy:

otcov odlet	otcovo zatknutie	otcova pozostalosť
otec odletel	otca zatkli	dedičstvo po otcovi
		dedičstvo patriace otcovi
otcova prísnosť	otcov nos	otcov kufor
otec je prísny	nos časť otca	kufor patriaci otcovi
otcova symfónia		otcova cesta
symfónia súvisiaca s otcom		otec cestuje
		otcova vozovka

Za prvé tri parafrázy je zodpovedné pravidlo 1. Toto pravidlo prijíma adjektíva, individuálne privlastňovacie, odvodené od substantív so sémantickým príznakom

osoba, stojace vľavo od substantíva, odvodeného od slovesa s nejakou intenčnou triedou. V tomto intenčnom poli sa potom vyhľadá tzv. *preferenčný pád*, t.j. prvý riadok slovesného rámca, ktorý unifikuje s daným sémantickým príznakom.

Najvyššiu aktuálnu aktiváciu v kontexte má objektová rola, preto LVAL hľadá objektový pád ako prvý. V intencii slovesa *zatýkať* naozaj nájde sémantický pád PACIENS, vyjadrený gramatickým akuzatívom.

V intencii slovesa *odletieť* sa nenájde žiaden objektový pád. Ďalšou najpreferovanejšou rolou je subjektová. V intenčnom rámci *odletieť* nachádzame AGENS vyjadrený nominatívom.

Pri slovese *zostať*$_2$ síce existuje subjektová rola, tá však neunifikuje s životnosťou *otca* (osoby nie sú predmetom dedičského konania). Ako preferenčný pád prichádza do úvahy prvý nasledujúci, a to SUKCESÍV. Vidíme, že je vyjadrený predložkovým pádom *po + lokál*. To, že ide o lexiu *zostať*$_2$ (konverz. zdediť), a nie o *zostať*$_1$ (syn. zotrvať), vyplýva z faktu, že *pozostalosť* sa nachádza v slovotvornom hniezde *zostať*$_2$, nie *zostať*$_1$.

Otcova pozostalosť však unifikuje zároveň aj s tretím pravidlom. Môže teda ísť aj o POSESÍV.

Štvrtá parafráza sa uskutočnila cez pravidlo č. 2, vďaka tomu, že *prísnosť* je deadjektívum.

Nasledujúce tri parafrázy umožnilo tretie pravidlo; s ohľadom na sémantické príznaky jednotlivých substantív. *Nos* je korporálium, dostáva pád PARTITÍV (časť celku). *Kufor* je konkrétne, ide o POSESÍV. *Symfónia* je abstraktum, preto zunifikuje iba s pádom KOHERENTÍV (neurčitá súvislosť). Otec mohol symfóniu zložiť, alebo môže byť jeho obľúbená. Konkretizovať tento pád je vyhradené neskoršej medzivetnej kontextovej analýze.

Posledný príklad demonštruje schopnosť riešiť polysémiu medzi *cesta*$_1$ (cestovanie) a *cesta*$_2$ (vozovka).

Uvedené parafrázy nie sú všetky, iba prvých niekoľko. Pád KOHERENTÍV prichádza do úvahy aj v niekoľkých ďalších prípadoch, ale jeho pravdepodobnostné ohodnotenie je nižšie v porovnaní s inými, dominantnejšími pádmi.

5.26 Kontrastívny príklad

náhrdelník z ametystu / koberec z Perzie / dar z lásky

hrdlo	(korp)	koberec	(vec)
náhrdelník	(vec)		
		Perzia	(lok)
ametyst	(mat)	perzský	(akvzt)
ametystový	(klas)		
		darovať	(303)
milovať	(100)	obdarovať	(304)
láska	(dej)	dar	(vec)

symbol(láska,srdce).

```
RVAL((sub,Se1),_,z,gen,(sub,Se2),_,SP)
    :- (sema(Se1,kon),Se2=mat,SP=mat);
       (Se2=lok,SP=ori).

RVAL((sub,_),(sl,_),z,gen,(sl,_),_,kaz).

LVAL((adj,akvzt,_,Sf),(sub,lok),(sub,Se,_,_),_,SP)
    :- (Sf=n;Sf=sk;Sf=ov;Sf=ick),
       ((Se=mov,SP=ori);(Se=nemov,SP=lok)).

LVAL((adj,klas,_,Sf),(sub,mat),(sub,Se,_,_),_,mat)
    :- (Sf='';Sf=ov;Sf=en;Sf=n),sema(Se,kon),sema(Se,neživ).
```

Pri uvedenej konfigurácii kolokačné funkcie určia pravostranným prívlastkom nasledujúce sémantické pády: v prvom prípade MATERIÁL, v druhom ORIGINATÍV (pôvod), v treťom KAUZATÍV (príčina). Dostávame tieto parafrázy:

náhrdelník z ametystu	koberec z Perzie	dar z lásky
ametystový náhrdelník	perzský koberec	dar zo srdca

Žiadne z pravidiel LVAL nedovolí parafrázovať *dar z lásky* ako *láskavý dar*, pretože KAUZATÍV sa takouto povrchovou formou neprejavuje. Ľavonominalizačné vzorce však umožnia formou desubstantívneho adjektíva vyjadriť MATERIÁL aj ORIGINATÍV.

5.27 Protipríklady

Senzitivita nominalizačných vzorcov LVAL a RVAL má svoje hranice. Napríklad fráza *mesačný modul* unifikuje s pravidlom č. 17 a dostane analýzu analogickú ako *indický koberec*:

modul
ORI_1» Mesiac

No *mesačný modul* nie je *modul z Mesiaca*, ale modul určený na výskum Mesiaca. Ani *perzský koberec* nemusí byť *koberec z Perzie*, ale môže mať iba perzský vzor.

Frázy *vodné čerpadlo, benzínové čerpadlo, liatinové čerpadlo* v našom sémantickom univerze úplne splynú. Dostanú rovnakú reprezentáciu, pretože *voda, benzín* aj *liatina* sú charakterizované jednou a tou istou sémantickou črtou *'materiálium'*. Ale prvé čerpadlo *čerpá vodu*, druhé je *poháňané benzínom* a tretie je *vyro-*

bené z liatiny. Ak by sme tieto tri prípady chceli rozlíšiť, museli by sme našu lexikálnu sémantiku rozšíriť o príznak *'palivá'*, t.j. materiály s latentným zdrojom energie, a zaviesť aj sémantické črty pre rôzne skupenstvá, aby sme mohli odlíšiť *'tuhé'* materiály od *'tekutých'*.

5.28 Súvetná sémantika

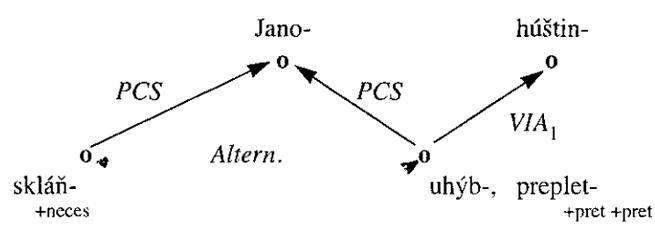
Sémantiku podraďovacích súvetí sme už prebrali: vedľajšie vety vystupujú vždy ako participanty slovesných alebo nominálnych fráz. Zostáva povedať niečo o analýze priraďovacích súvetí. Analýzou priraďovacieho súvetia rozumieme:

a) Analýzu každej vety súvetia osobitne;

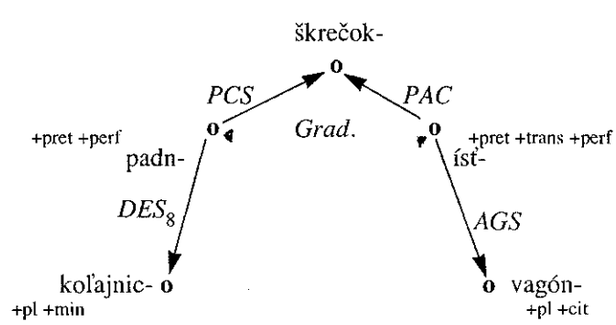
b) Určenie vzťahu medzi týmito vetami. Konkrétne to znamená spojiť ich sieťové reprezentácie hranou, označenou niektorým z 16-tich symbolov možných typov súvetia (pozri prehľad v 1.3).

5.29 Príklady

Jedno zlučovacie súvetie so striedavými dejmi a jedno gradačné súvetie:



Tu sa musel skláňať, tam uhýbať, prepletať húštinou. (Horák)



Škrečok spadol na koľajničky a ešte ho prešli aj vagóniky.

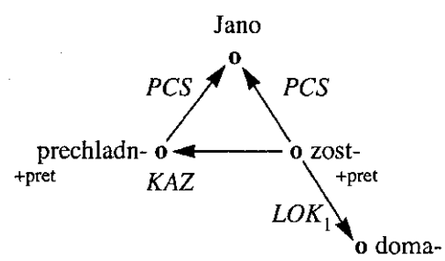
5.30 Protipríklady

Identifikačným znamením každého súvetia sú spojky. Avšak niektoré spojky môžu uvádzať viac ako jeden typ súvetia. Napríklad spojka *a* môže fungovať ako zlučovacia, ale aj ako odporovacia. Rozlíšenie je dané iba obsahom viet. V týchto prípadoch SAPFO podáva dvojznačnú a aj nesprávnu analýzu.

Osobitnými typmi priraďovacích súvetí, v ktorých sa forma nezhoduje s obsahom, sú *nepravá parataxa* a *nepravá hypotaxa*. SAPFO ich nedokáže identifikovať.

5.31 Nepravá parataxa

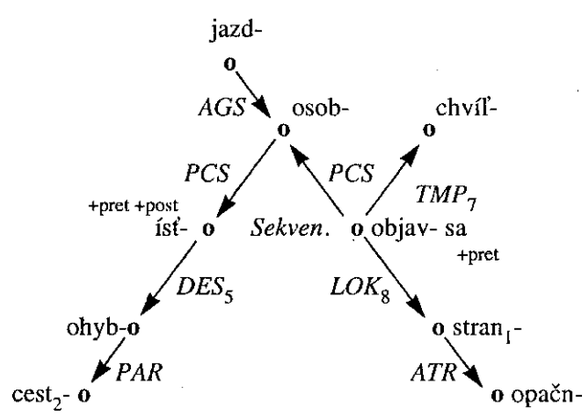
Formálne je to priraďovacie súvetie, hoci významový vzťah medzi vetami je podraďovací:



Je prechladnutý, zostal doma.

5.32 Nepravá hypotaxa

Opačný prípad. Ide o súvetie, ktoré je podraďovacie svojou formou, ale priraďovacie významom:



Jazdec zašiel za ohyb cesty, aby sa po chvíli vynoril na opačnej strane.

5.33 Sémantika infinitívnych fráz

Slovesným aktantom môže byť aj infinitívna fráza. Infinitív môžu priberať niektoré druhy slovies (dicendi, percipiendi, cognoscendi). Takáto možnosť je vyznačená v intenčnom rámci slovesa pod druhou čiarou. Infinitív dostane rolu, ktorá je preň v intenčnom rámci vyznačená. Vnútorná štruktúra infinitívnej frázy je analogická slovesnej fráze - opäť ide o centrálne sloveso obklopené participantmi - a rieši sa rekurzívnym volaním funkcie VAL. Ak vo valenčnom systéme infinitívu chýba agensový aktant, stáva sa ním posledne uvedený aktant centrálneho slovesa, ktorý tak vystupuje v dvojakej sémantickej role, na obe strany, ako súčasť oboch slovesných valenčných systémov:

Peter sa rozhodol postaviť dom.

rozhodov- sa $^{+perf\ +pret}$
KOG» Peter-
FEN» stav- $^{+perf}$
AGS» = KOG
PAC» dom-

prikazov- $^{+perf\ +pret}$
EDI» Peter-
ADS» Pavol-
TEM» stav- $^{+perf}$
AGS» = ADS
PAC» dom-

Peter prikázal Pavlovi postaviť dom.

5.34 Sémantika adjektiválnych fráz

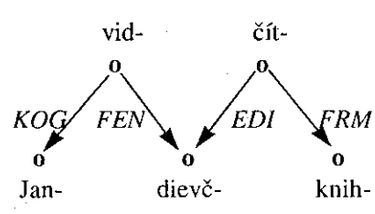
Nominálna fráza môže byť rozvitá adjektiválnou frázou. Adjektiválna fráza sa napája na hlavné substantívum prostredníctvom príčastí (trpného a činného minulého či prítomného). Predstavuje druhotnú predikáciu k substantívu, ktoré určuje.

Vnútorná štruktúra adjektiválnej frázy je znova analogická vete - centrálne sloveso v tvare príčastia je obklopené participantmi. Jedným z participantov je určované substantívum, ktoré tak participuje v dvoch dejoch, v dvoch valenčných systémoch súčasne. V každom z nich môže vystupovať v odlišnej sémantickej role.

Adjektiválne frázy reprezentujeme tak, akoby obe predikácie boli rovnocenné:

Jano vidí dievča čítajúce knihu.

vid-
KOG» Jano-
FEN» dievč-
«EDI čít-
FRM» knih-



To nám neskôr umožní tvoriť symetrické parafrázy typu:

Jano vidí dievča čítajúce knihu.
Janom videné dievča číta knihu.

5.35 Translokácia vetných členov

Všetky doteraz uvedené algoritmy sémantickej interpretácie mali vlastnosť *lokálnosti*; t. j. platilo, že dva členy boli v sémantickom vzťahu práve vtedy, keď boli aj vo vzťahu syntaktickom. Preto stačilo brať syntagmy jednu po druhej. Každú dvojicu 'nadradený člen - podradený člen' stačilo skúmať osobitne. Ak boli dva členy vo vzťahu syntaktickej podradenosti, boli podradené aj sémanticky a tento vzťah sa identifikoval špecificky ako jeden zo sémantických pádov.

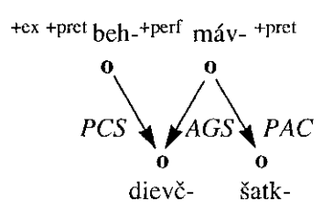
Existujú však prípady, keď vetný člen je formálne podradený prísudku, ale významovo určuje (doplňuje) podmet, predmet alebo iný vetný člen. Túto vlastnosť majú *prechodník, doplnok, absolútny nominatív, absolútny akuzatív a niektoré ďalšie parentetické útvary.*

Zhodné doplnky vyjadrujú svoju príslušnosť k podmetu (predmetu) aspoň zhodou v gramatických kategóriách. Ale nezhodný doplnok sa k vetnému členu, ku ktorému patrí, neviaže ani zhodou, ani väzbou, ani primkýnaním a stojí osamostatnený úplne mimo nominálnej frázy, do ktorej by podľa významu patril.

Syntaktický analyzátor postupuje vo všeobecnosti tak, že analyzuje aktanty okolo slovesa jeden po druhom. Ak narazí na prechodníkovú alebo doplnkovú konštrukciu, mal by ju v sémantickej sieti pripojiť k agensu (resp. paciensu). Ale analýza agensa/paciensa je buď už hotová (ak vo vete predchádzal), alebo sa ešte nezačala (ak vo vete ešte len má nasledovať). SAPFO používa zvláštny register pre vzdialené závislosti **far**(), do ktorého dočasne ukladá polopredikácie. Až vo chvíli, keď sa objavia oba zúčastnené členy - určujúci aj určovaný - polopredikácia sa vytiahne z pomocného registra a pripojí sa do sémantickej siete k agensu (resp. paciensu):

Dievča vybehlo mávajúc šatkou.

bež- +ex +perf +pret
PCS» dievč-
«AGS máv-
PAC» šatk-



Rovnako ako pri adjektiválnych frázach, naša reprezentácia zodpovedá dvom samostatným vetám:

Dievča vybehlo. Dievča máva šatkou.

5.36 Sémantika doplnku

Doplnok (druhotná predikácia) môže mať formu prechodníka, príčastia, adjektíva, radovej číslovky, infinitívu, vedľajšej vety so spojkou *ako*, a tiež formu rôznych predložkových fráz. Vo všetkých prípadoch analyzátor pripája druhotnú predikáciu k doplňovanému vetnému členu:

Ivan prišiel prvý.

ísť-$^{+ad\ +pret}$
PCS» Ivan-
ORD» jeden-

Jano vidí rásť fialku.

vid-
KOG» Jan_1-
FEN» fialk-
«PCS rásť-

Jana sedela zamyslená.

sed-$^{+pret}$
PCS» Jan-
«KOG zamýšľ- sa^{+perf}

Videl Oľgine oči, ako sa smiali.

vid-$^{+pret}$
FEN» ok-$^{+pl}$
PAR» Oľg-
«PCS smiať- sa^{+pret}

Jadro horeuvedených doplnkov je slovesné. Funkcia VAL ich pripojila k AGENSu tak, že našla jeho rolu, ktorú by mal vo valenčnom systéme doplnkového slovesa.

Čisto adjektívny doplnok sa pripája k AGENSu prostredníctvom pádu ATRIBÚT:

Jana sa vrátila šťastná.

vrát- sa^{+pret}
PCS» Jan-
ATR» šťastn-

Stavový nominatív a akuzatív sa pripájajú ešte inak. Hlavný člen majú nominálny a k AGENSu sa pripájajú akoby bol ich posesívnym prívlastkom. (*Jano sedí, lakte opreté o kolená = Jano sedí. Janove lakte sú opreté o kolená.*)

Funkcia LVAL teda absolútnym pádom priradí príslušnú rolu, v závislosti od ich sémantických príznakov. Najčastejšie pôjde o POSESÍV (kon.) alebo PARTITÍV (korp.):

Jano sedí, lakte opreté o kolená.

sed-
PCS» Jan_1-
«PAR laket'-$^{+pl}$
«PAC oprieť-
DES» kolen-$^{+pl}$

kráč-
PCS» Ivan
«POS sieťovk-
«PAC pln-$^{+perf}$
REA» šalát-

Ivan kráča, sieťovka plná šalátu.

Z doplnkov vyjadrených predložkovými pádmi sú najčastejšie *s + inštrumentál, v + lokál, bez + genitív, pri + lokál*. Doplnok je kondenzovaná, pôvodom samostatná druhotná predikácia pripojená v nominalizovanej forme. Vetu s doplnkom možno vždy rozložiť na dve samostatné vety. Sémantickú rolu doplnku určujú funkcie VAL, LVAL, RVAL podľa toho, či ide o doplnok deverbatívny, deadjektívny alebo desubstantívny:

Jana sa dívala s rozochvenými perami.

dív- sa^{+pret}
KOG» Jan$_2$-
«PAR per-$^{+pl}$
«PAC chvieť- sa$^{+perf+inc}$

1. Jana sa dívala.
2. Janine pery boli rozochvené.

Jana sa dívala s pohľadom plným lásky.

dív- sa^{+pret}
KOG» Jan$_2$-
«KOG hľad-$^{+pret}$
«PAC pln-
REA» lásk-

1. Jana sa dívala.
2. Hľadela láskyplne.

Jano prišiel domov v dobrej nálade.

ísť-$^{+ad\ +perf}$
DES$_1$» domov
PCS» Jan$_1$
«AFE nalad- sa
ATR» dobr-

1. Jano prišiel domov.
2. Jano bol dobre naladený.

Vo forme predložkových pádov sa doplnok stretáva s príslovkovým určením okolnosti. Takýto doplnok plní dvojakú funkciu: určuje aj podmet (predmet), aj prísudok zároveň (Kačala, 1971). Má teda dva nadradené vetné členy! Riešenie ambivalencie doplnku uvedieme v odseku 12.10.

5.37 Kontrastívny príklad

Všimnite si rozdiel v analýze a reprezentácii týchto dvoch syntakticky rovnorodých viet:

Kráľ vyzval kapitána viesť voj.

vyzv-$^{+pret}$
EDI» kráľ-
ADS» kapitán-
TEM» viesť-
AGS» = ADS
PAC» voj-

Kráľ videl kapitána viesť voj.

vid-$^{+pret}$
KOG» kráľ-
FEN» kapitán-
«AGS viesť-
PAC» voj

V prvom prípade infinitívna konštrukcia iba realizuje jeden z povinných aktantov hlavného slovesa. V druhom prípade je infinitív doplnkom. Druhá veta sa dá parafrázovať dvoma samostatnými vetami: *Kráľ videl kapitána. Kapitán viedol voj.* Ide o dva súbežné, ale samostatné deje. V prvom prípade ide o dej jeden a veta sa dá parafrázovať iba vcelku: *Kráľ vyzval kapitána, aby viedol voj. Kráľ vyzval kapitána na vedenie voja.*

5.38 Združené pomenovania

Osobitný problém pre automatickú syntaktickú analýzu predstavujú združené pomenovania (*vysoká škola, podstatné meno, vlčí mak, pastierska kapsička, žabie očko)*. Tieto sa nedajú identifikovať formálne, ale iba slovníkovo.

Zdanlivo by sa táto úloha dala riešiť polysemickou diferenciáciou. Zaviedli by sme napr. ešte jeden význam adjektíva *vysoký*. To je však v spore s tým, že táto nová adjektívna lexia by sa nedokázala spájať so substantívami všeobecne, ale iba s jedným jediným - s koreňom *škol-*, a tiež by sa nedokázala zúčastňovať syntaktických transpozícií.

SAPFO rieši združené pomenovania tak, že sú vymenované v lexikálno-sémantickom slovníku, kde sú už predspracované, takže nepodliehajú regulárnej syntaxi:

$$\text{združ}\left(\begin{bmatrix}\text{škol-} \\ \text{LDET: vysok-}\end{bmatrix}\right).$$

Sémantický pád medzi členmi združeného pomenovania (pozri aj 9.1) sa neidentifikuje, pretože pomenovanie sa chápe ako celok, ako jedna lexéma. Tým sa takýto kus siete stáva aj sémanticky petrifikovaný, pretože neunifikuje so žiadnou z valenčno-transpozičných funkcií (VAL, LVAL, RVAL) a nemôže sa zúčastňovať sémantických transformácií a syntaktických transpozícií. To je však presne to, čo od združeného pomenovania očakávame. Inak by vznikali nežiadúce parafrázy:

vysoká škola $^{*}\!\rightarrow$ výška školy
podstatné meno $^{*}\!\rightarrow$ meno s podstatou
slepá ulička $^{*}\!\rightarrow$ ulička, ktorá nevidí
vlčí mak $^{*}\!\rightarrow$ mak zvierat

Teda na jednej strane tu máme viacslovné pomenovania, ktoré chápeme ako jednoliaty významový celok, pretože majú ustálenú formu *(slepá ulička)*, na druhej strane jednoslovné kompozitá, ktoré štiepime na dva aj viac významových aktantov, pretože dokážu fungovať samostatne (*maloobchodník, veľkopoľnohospodárstvo)*. To je v súlade s princípom diferenciácie významového kontinua na základe kontrastujúcich foriem. Podľa toho frázu *oxid uhličitý* už nepokladáme za združené pomenovanie, pretože fráza *oxidy uhlíka* je jej parafrázou, ktorá vznikla transpozíciou a ktorá s ňou kontrastuje.

Analyzátor je naostatok neúspešný v tých prípadoch, keď ide zdanlivo o združené pomenovanie, ktoré ním ale v skutočnosti nie je. Vtedy bude analýza chybná alebo zostane dvojznačná, napr.:

Chodím do tej vysokej školy. (Ukazujúc na ZŠ.)

5.39 Modálna stavba vety

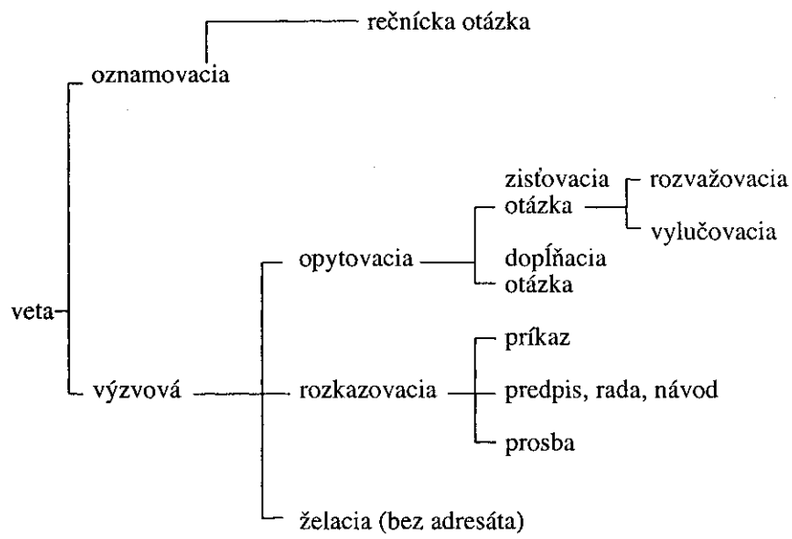
K úplnej sémantickej analýze patrí aj identifikácia modálnej stavby vety.

Postojová modálnosť vyjadruje komunikatívny zámer hovoriaceho. Identifikačným znamením *oznamovacej* vety je bodka, prísudok v indikatíve alebo kondicionáli a klesavá kadencia. *Rozkazovací* spôsob sa pozná podľa prísudku v imperatíve, výkričníka a uvádzacích pobádacích častíc. *Želacie* vety charakterizujú najmä želacie častice a podmieňovací spôsob. *Otázky* sa končia otáznikom a stúpavou kadenciou. V dopĺňacích otázkach sú opytovacie zámená, zisťovacie obsahujú opytovacie častice a inverziu podmetu a prísudku.

Istotnou modálnosťou hovoriaci vyjadruje istotu o stupni pravdivosti svojej výpovede. Môže byť vyznačená futúrom alebo kondicionálom slovesa, no hlavne hodnotiacimi časticami na osi áno - nie (*možno, akiste, čertvie...)*.

Voluntatívna modálnosť vyjadruje možnosť, potrebu, nevyhnutnosť, schopnosť, vôľu, sklon. Vyjadruje sa modálnymi a polomodálnymi slovesami (*musieť, chcieť, môcť*...) alebo modálnymi príslovkami *((ne)treba, (ne)možno, (ne)hodno, (ne)slobodno...)*.

Postojová a voluntatívna modálnosť pripájajú k výslednej analýze vety po jednom komponente významu (pozri zoznam v 11.3). Istotná modálnosť sa vyjadrí číslom intervalu <0,1> (rozhodne = 1.00, zaiste = 0.90, zrejme = 0.75, snáď = 0.30, horkýže = 0.05).



Delenie viet podľa postojovej modálnosti.

Uvedené formálne identifikačné znaky nestačia vždy na správne určenie vetnej modálnosti. M. A. K. Halliday (1976) upozorňuje, že jazyk je sociálny proces, v ktorom každý účastník musí vedieť o cieľoch, presvedčeniach a sociálnej role druhého účastníka v danom kontexte, situácii. Napríklad otcova veta adresovaná synovi *Dobrí chlapci také veci nerobia* je skôr výzvou ako oznamovacou vetou. Takisto veta *Snažím sa zatĺcť tento klinec* vyslovená v určitej situácii nie je oznamom, ale prosbou o pomoc.

Osobitným druhom oznámenia je rečnícka otázka. Má formu opytovacej vety, avšak odpoveď je už obsiahnutá v otázke. SAPFO nedokáže identifikovať rečnícke otázky ani žiadne iné druhy viet, ktorých komunikačný zámer viditeľne nezodpovedá ich forme. sú SAPFO nedokáže identifikovať ani gnómickú osobu (*Nechváľ deň pred západom slnka. Povedz mi s kým chodíš, a ja ti poviem, kto si*), autorský plurál (*Podarilo sa nám dokázať, že...*).

Plurál majestaticus *(My, kráľ z vôľe Božej,...)* a familiárny plurál (*Ako sa máme?)* sa prezradia iba vtedy, ak za nimi nasleduje menná fráza, ktorá svojím jednotným číslom plurálu protirečí. SAPFO nerozozná ani tretiu osobu vo funkcii druhej osoby (*Slečna sa na mňa hnevá?)*

Ďalej nedokáže rozoznať gnómický čas (*Lož má krátke nohy. Drevo sa píli)*, historický prézens (*Sokrates hovorí...*), prézens v zastúpení futúra (*zajtra maľujeme*), budúci čas vo funkcii odhadu (*vonku bude asi poriadna zima)*, gnómický imperatív *(Sprav čertovi dobre, peklom sa ti odslúži)* a minulý čas ako zosilený imperatív *(Čert ho vzal)*.

Nie vždy dokáže rozoznať indikatív a kondicionál vo funkcii imperatívu (*Ideš sem! Platím! Zajtra ideš k lekárovi. Keby už bol mier na svete!)* a želací infinitív (*Tak odísť niekam ďaleko!)*

6.

INTEGROVANÁ SYNTAKTICKO-SÉMANTICKÁ ANALÝZA

V tejto kapitole ukážeme, že oddeľovanie jednotlivých jazykových procesov (úrovní) nie je prirodzené, a že pri striktne sekvenčnej analýze jazyka je jedným z hlavných faktorov zodpovedných za exponenciálny nárast počtu nejednoznačností. Ukážeme, že obdivuhodná schopnosť ľudskej mysle vnímať jazyk je možná okrem iného aj vďaka tomu, že človek vníma jazyk ako celok, a že takéto paralelné spracovanie jazyka je možné, špeciálne vo flektívnych jazykoch, úspešne modelovať na počítači. Syntaktický a sémantický modul nemusia pracovať osamotene. Naopak. Možnosť permanentnej komunikácie medzi nimi nám dovolí podstatným spôsobom znížiť počet viacznačností a dodať celej analýze na efektivite a elegancii.

6.1 Nárast nejednoznačností pri sekvenčnej analýze

Za exponenciálny vzrast počtu nejednoznačností počas syntaktickej analýzy sú zodpovedné predovšetkým dva faktory: lexikálna homonymia (polysémia) a syntakticko-štrukturálne viacznačnosti. Napríklad pri analýze mennej frázy

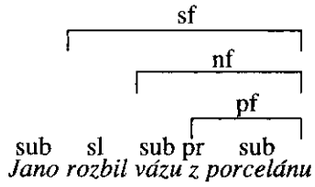
medený vodič

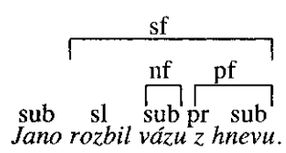
treba rozhodnúť, o ktorý z dvoch *vodičov*, prítomných v slovníku, ide:

vodič$_1$ (+živ)
vodič$_2$ (+inštr)

Také rozhodnutie však spadá mimo rámca morfológie a syntaxe. Tieto musia pracovať s oboma alternatívami ako s možnými, až do času, kým pravidlá sémantickej spájateľnosti nerozhodnú v prospech *vodiča*$_2$. Už výsledok morfologickej analýzy bude dvojznačný, lebo životný *vodič*$_1$ sa skloňuje inak ako neživotný *vodič*$_2$.

Iný príklad. Nasledujúce dve vety majú identickú povrchovú štruktúru, ale ich syntaktické stromy sa líšia:





Je nemožné, aby čistá syntax dokázala zistiť, že predložková fráza na konci vety je v prvom prípade súčasťou mennej frázy *váza z porcelánu* a v druhom prípade patrí priamo pod centrálne sloveso *rozbiť z hnevu*. Také rozhodnutie možno vykonať iba s ohľadom na lexikálnu sémantiku slov vo vete:

porcelán	(+mat)
hnev	(+abstr)

Pri uvažovaní tejto frázy by syntax potrebovala pomoc sémantiky. Tá však odštartuje svoju analýzu až po ukončení syntaxe a preto nemôže nijako pomôcť. Teda syntax musí s touto dvojznačnosťou (ktorá násobí počet všetkých ostatných nejednoznačností vo vete dvoma) pracovať až do ukončenia syntaktickej analýzy a nakoniec ju odovzdať sémantickému modulu nerozhodnutú.

6.2 Paralelná syntakticko-sémantická analýza

Sekvenčne pracujúce moduly si teda posúvajú viacznačnosti a nezmyselné interpretácie medzi sebou. Možnosť vzájomnej komunikácie by bola veľmi žiadúca. Zdá sa, že človek ani nedokáže vnímať syntaktickú štruktúru vety bez toho, aby si súčasne neuvedomoval lexikálnu sémantiku slov. Preto ho gramaticky prípustné, ale sémanticky nezmyselné interpretácie ani nenapadajú a nezdržujú.

Hypotézu, že ľudská myseľ spracúva jazyk ako celok a že syntax a sémantika u človeka fungujú ako dva samostatné, ale úzko komunikujúce paralelné procesy, sme sa pokúsili modelovať v našom počítačovom modeli. Integrovaný syntakticko-sémantický analyzátor SAPFO sa nezastavuje pri priraďovaní syntaktických rôl. Nekonštruuje ani žiaden syntaktický strom, ale snaží sa priradiť sémantickú rolu každej podfráze vždy bez meškania, ihneď ako na ňu narazí. To vôbec neznamená, že dôležitosť syntaxe by bola nejakým spôsobom oslabená, ako je to v niektorých sémantických gramatikách. Syntaktický analyzátor iba neodkladá konzultáciu so sémantickým modulom a jeho znalosťami. Kedykoľvek syntax identifikuje nejakú podfrázu, napríklad mennú frázu, zavolá sémantický modul (kolokačné funkcie VAL, LVAL, RVAL), aby identifikáciou jej sémantickej roly potvrdili aj jej sémantickú prípustnosť. Ak sémantika zistí, že podfráza je síce syntakticky v poriadku, ale sémanticky je nezmyselná, upozorní syntax, že pokračovať v práci s touto, povedzme mennou, frázou je neperspektívne. Vylučujúc nezmyselné syntagmy už v ranom štádiu spracovania vety môže sémantika ušetriť syntaxi množstvo zbytočnej práce. Napríklad veta

Dub takýchto rozmerov váha za žiadnych okolností nemôže odvážiť.

zmätie a do slepej uličky zavedie každý syntaktický analyzátor zhora-nadol a zľava-doprava tým, že *dub* vyzerá ako podmet a *váhať* ako hlavné sloveso (*Dub váha...*), až kým sa nepríde po ôsme slovo. Tam potom analyzátor zlyhá.

Integrovaná syntax-sémantika SAPFO zachráni túto situáciu už pri prvom slove. Podľa intenčných rámcov slovies *váhať* a *odvážiť*

VAL(váhať,0,nom,živ,KOG).	VAL(odvážiť,0,nom,živ,AGS).
VAL(váhať,nad,ins,_,FEN).	VAL(odvážiť,0,nom,inštr,INS).
VAL(váhať,kvôli,dat,_,KAZ).	VAL(odvážiť,0,dat,živ,BEN).
	VAL(odvážiť,0,dat,[kon,než],PAR).
	VAL(odvážiť,0,ak,kon,PAC).
	VAL(odvážiť,0,ins,inštr,INS).
	VAL(odvážiť,na,lok,inštr,INS).
	VAL(odvážiť,pre,ak,_,MOT).

neživotný *dub* v intenčnom poli slovesa *váhať* vôbec nemôže stáť. Aby sa *dub* mohol stať kognitívnym subjektom procesu *váhania*, musel by byť životný, pretože *váhať* pripúšťa iba životný subjekt a priamy objekt nepripúšťa vôbec. Neživotné substantívum *dub* teda nevyhovuje či už je v prvom alebo v štvrtom páde. V intenčnom poli slovesa *odvážiť* sa však bez ťažkostí uplatní ako akuzatívny PACIENS.

Aj prípustnosť fráz *váza z porcelánu, rozbiť z porcelánu, váza z hnevu, rozbiť z hnevu* možno identifikovať okamžite. Prvá a štvrtá dostanú svoje sémantické pády, zatiaľčo druhá a tretia skončia neúspechom:

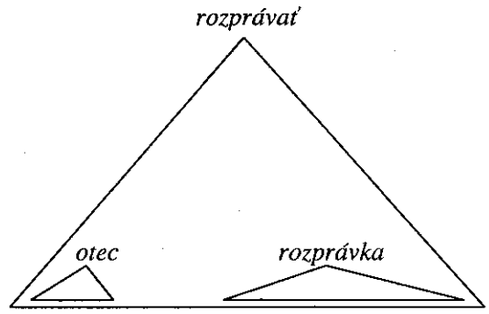
RVAL(váza,_,z,gen,porcelán,_) = MATERIÁL
VAL(rozbiť,_,z,gen,porcelán,_) = FAIL
RVAL(váza,_,z,gen,hnev,hnevať_sa) = FAIL
VAL(rozbiť,_,z,gen,hnev,hnevať_sa) = KAUZATÍV

Jeden z ľavovalenčných nominalizačných vzorcov umožní pripojiť *meď* k neživotnému *vodiču*$_2$ ako MATERIÁL. Ale spojenie životného *vodiča*$_1$ s dematerialnym adjektívom *medený* žiadna kolokačná funkcia neumožní:

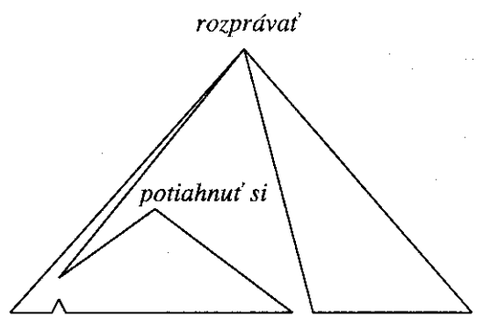
LVAL(medený,meď,vodič$_2$,viesť$_2$) = MATERIÁL
LVAL(medený,meď,vodič$_1$,viesť$_1$) = FAIL

6.3 Predvýpočet valenčných polí

Pod *valenčným poľom* vetného člena rozumieme tú časť vety, v ktorej sa nachádzajú aktanty sémanticky podradené tomuto členu. Znázornenie:



Starý otec rozpráva rozprávku o Snehulienke.



Ivan potiahnuc si z cigarety začal rozprávať.

Videli sme, že priraďovanie sémantických rôl frázam bezprostredne po ich identifikácii malo viaceré výhody. V čom je teda problém? Aby sme mohli aktantom pomocou funkcií VAL, LVAL, RVAL priraďovať ich sémantické roly, musíme najprv vedieť, v akom valenčnom poli sa nachádzajú. Musíme teda poznať centrálny vetný člen. Potrebujeme ho zadať ako argument do kolokačných valenčných funkcií. Pri LR-syntaktickej analýze sa však aktanty umiestnené naľavo od centrálneho slovesa (resp. substantíva) objavujú skôr ako ich nadradený člen. Hlavné sloveso tiež niekedy môže stáť až na konci vety (ako v nemčine a v latinčine).

Ak teda chceme aktantom priraďovať ich sémantické roly okamžite, musíme najprv zistiť, v akom valenčnom poli sa nachádzajú. Potom môžeme odštartovať analýzu s účinnou sémantickou výzbrojou.

SAPFO identifikuje valenčné polia jednoduchým predvýpočtom:

Algoritmus na identifikáciu centrálneho slovesa:

(1) Nájdi prvé nemodálne, nefázové určité sloveso alebo neurčitok, ignorujúc všetko medzi dvoma čiarkami.

Algoritmus na identifikáciu centrálneho substantíva:

(1) Nájdi najľavejšie podstatné meno.

(2) Ak za ním nasleduje podstatné meno v rovnakom páde, druhé meno je hlavné, prvé je prístavok.

(3) Ak nasleduje slovesné príčastie, musia sa zhodovať v rode, čísle a páde, inak pokračuj v hľadaní.

Napríklad nasledujúca menná fráza obsahuje tri podstatné mená. Podľa uvedenej procedúry je centrálnym to posledné:

slnkom zaliate mesto ***Bratislava***

Tieto algoritmy dávajú spravidla jednoznačný výsledok. Viacznačné sú iba v prípade, keď sa vo vete objavia vedľajšie vety a koordinácie s nepárnym počtom čiarok. Vtedy nastane zmätok o tom, či sa práve nachádzame v hlavnej alebo vedľajšej vete a sloveso z vedľajšej vety sa omylom tiež považuje za kandidáta na hlavné sloveso:

Peter, Pavol a Quido, ktorí nevedeli, o čo ide, sa smiali.

6.4 Špecifikum flektívnych jazykov

Práve pri predvýpočte valenčných polí vychádza najavo rozdiel medzi slovenčinou a angličtinou, ktorý je z hľadiska výpočtovej zložitosti podstatný. Spomeňme si, čo sme povedali o flektívnom charaktere slovenčiny. V slovenčine je skoro nemožné zameniť sloveso za iný slovný druh. Príklad s tvarovou homonymiou *Dub váha odváži* je v angličtine frekventovaný, ale pre slovenčinu je netypický. Slovenské sloveso sa skoro vždy dá identifikovať bez uvažovania syntaktického kontextu. Špecifické morfologické príznaky (napr. sufixy *-ť, -al, -il)* ho viditeľne odlišujú od ostatných slovných druhov. „Odhaliť" slovenské sloveso nie je žiaden skutočný problém.

Za cenu veľmi jednoduchého predvýpočtu (simple pattern matching) takto získavame veľkú výhodu integrácie syntaxe a sémantiky do paralelnej procedúry. To však platí iba pre flektívne jazyky. V angličtine o väčšine slov nemožno povedať, či ide o sloveso alebo iný slovný druh, keď stoja samostatne. Anglické sloveso (resp. substantívum) sa konštituuje až svojou pozíciou a vzťahom k ostatným slovám vo vete. Z toho vyplýva, že vykonať predvýpočet na identifikáciu centrálneho vetného člena v angličtine by znamenalo vykonať bezmála kompletnú syntaktickú analýzu - a preto ho za „predvýpočet" k syntaktickej analýze vôbec nemožno pokladať!

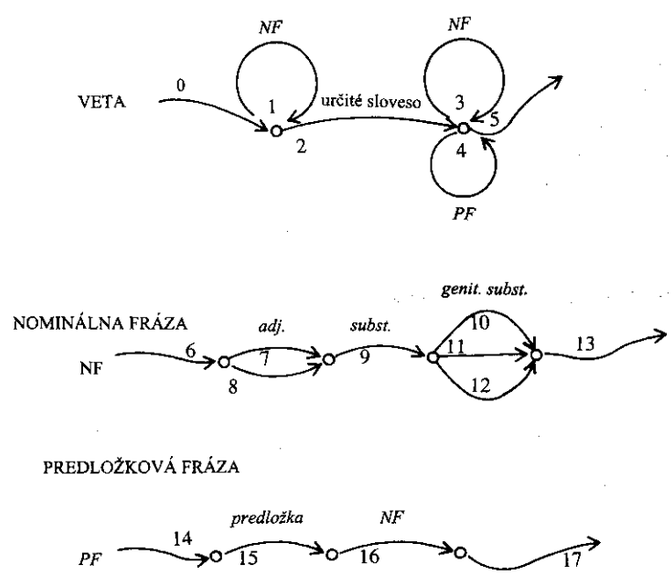
6.5 Interakcia syntaxe, sémantiky a morfológie. Príklad

Spoluprácu syntaxe, sémantiky a morfológie na pozadí slovníka a sémantických valenčných rovníc ilustrujme na príklade typickom pre slovenčinu. Pre anglický

analyzátor je typický backtracking, nejednoznačnosti, slepé uličky. Slovenský analyzátor je zdanlivo zložitejší, ale napreduje viac deterministicky, opierajúc sa o bohatý pádový a tvarový systém. Analyzujme vetu

Južné krídlo hradu postavili Florentskí majstri z guľatých skál.

pomocou ATN-siete



hrany	podmienka + akcia
1,2,3	Zhoda subjektu (menná fráza v nominatíve) so slovesom v osobe, čísle a životnosti.
7,9	Zhoda prídavného mena s podstatným menom v páde, čísle, rode a životnosti.
5	Hlava automatu na konci vstupného reťazca.
1,3,4	Funkcia VAL sa pokúsi priradiť sémantické roly participantom vo forme nominálnej alebo predložkovej frázy.
7	Funkcia LVAL priradí rolu ľavostrannému prívlastku hlavného mena mennej frázy.
10,12	Funkcia RVAL priradí sémantickú rolu pravostrannému prívlastku v mennej fráze.

Použité valenčné rovnice:

intenčný rámec manuálnych konštruktívnych procesov [20]:
(postaviť)

```
VAL(0,nom,živ,ags).
VAL(0,nom,inštr,ins).
VAL(0,ak,konkr,pac).
VAL(0,dat,živ,ben).
VAL(0,inš,inštr,ins).
VAL(0,inš,abstr,mds).
VAL(z,gen,mat,mat).
VAL(do,gen,mat,mat).
VAL(na,lok,inštr,ins).
VAL(na,ak,_,mot).
VAL(pre,ak,živ,ben).
VAL(za,ak,živ,sub).
VAL(za,gen,prír,ele).
```

(1) LVAL((adj,rel,_,-ný),(sub,lok),(sub,nemov,_,_),_) = LOK
(2) LVAL((adj,rel,_,-ný),(sub,lok),(sub,mov,_,_),_) = ORI
(3) LVAL((adj,_,_,-ský),(sub,lok),(sub,mov,_,_),_) = ORI
(4) LVAL((adj,rel,_,-atý),(sub,_),km(sub,konkr,_,_),_)= REL

(5) RVAL((sub,mov),_,0,gen,(sub,mov),_) = PAR
(6) RVAL((sub,nemov),_,0,gen,(sub,nemov),_) = PAR
(7) RVAL((sub,neživ-konkr),_,z,gen,(sub,mat),_) = MAT

(8) DVAL((sl,I-dok,_,po-,K1),(sl,I-nedok,_,_,K2)) = [K2^{+perf}]

Nech morfologická analýza dáva takýto výstup:

južné relačné adjektívum *južný*, rod ženský alebo stredný, pád nominatív alebo akuzatív, číslo singulár alebo plurál, životné alebo neživotné, odvodené od substantíva *juh*, lokália, sufixom *-ný*.

krídlo substantívum *krídlo*, rod stredný, pád nominatív alebo akuzatív, číslo singulár, neživotné, nepremiestňovateľné (krídlo hradu) alebo premiestňovateľné (vtáčie krídlo).

hradu substantívum *hrad*, rod mužský, pád genitív alebo datív, číslo singulár, neživotné, nehnuteľné.

postavili l-príčastie od *postaviť*, 3.osoba, číslo plurál, minulý čas, zápor pozitív, intenčný rámec č.20.

Florentskí adjektívum relačné *Florentský*, rod mužský, pád nominatív, číslo plurál, životné; odvodené od substantíva *Florencia*, lokália, sufixom *-ský*.

majstri substantívum *majster*, rod mužský, pád nominatív, číslo plurál, životné, neodvodené.

z genitívna predložka

guľatých adjektívum *guľatý*, rod mužský, ženský alebo stredný, pád genitív alebo lokál, číslo plurál, životné alebo neživotné, desubstantívne od *guľa*, sufixom *-atý*.

skál substantívum *skala*, rod ženský, pád genitív, číslo plurál, neživotné, neodvodené.

Začnime analýzu. Všimnite si, že aj keď je morfologická analýza čo do gramatických kategórií viacznačná, žiadne slovo nespadá do viac ako jedného slovného druhu súčasne. Identifikovať hlavný vetný člen predvýpočtom nebude problém. Najľavejšie nemodálne sloveso - *postaviť* - je centrálne. Použitím derivačného pra-

vidla (8) ho rozložíme na ireducibilný koreň a komponenty významu; a začneme budovať zlomky sémantickej siete:

MANUÁLNY_KONŠTRUKTÍVNY_PROCES: *stav-*$^{+pret\ +perf}$

Veta sa nezačína slovesom (hrana 2). Musíme zanalyzovať mennú frázu (hrana 1). Horeuvedeným predvýpočtom určíme najprv hlavu mennej frázy - *krídlo*. Gramatické kategórie slova *južný* a *krídlo* zunifikujú do: rod stredný, singulár, nominatív alebo akuzatív. Sémantická rola slova *južný* (podľa 1. a 2. ľavovalenčného pravidla) závisí od premiestňovateľnosti slova *krídlo*. Ale *krídlo* je polysemické slovo, môže byť premiestňovateľné aj nepremiestňovateľné. Analyzátor nevie, či *juh* je miestom (kde) hradného krídla, alebo jeho pôvodom (odkiaľ), ako napr. vo fráze *južné ovocie*. Musí zatiaľ uvažovať obe alternatívy.

Ďalšie slovo je *hradu*, v datíve alebo genitíve. Mohli by sme opustiť NF-sieť po hrane č.11 a snažiť sa identifikovať *hradu* ako datívový objekt (ako ďalšiu nf na hrane 1). Ale štvrtý riadok slovesného valenčného rámca tomu zabráni, zlyhajúc na požiadavke životnosti BENEFAKTORa. Preto *hradu* je v genitíve, súc pravostranným prívlastkom slova *krídlo* (hrana 10). Prvé dve pravostranné valenčné pravidlá (5), (6) stelesňujú poznatok, že dva objekty, ak jeden je časťou druhého, sú buď oba premiestňovateľné, alebo oba nepremiestňovateľné. Tak nepremiestňovateľnosť *hradu* spätne implikuje nepremiestňovateľnosť *krídla* a toto rozhodne o role adjektíva *južný* ako o LOKATÍVe. Zanalyzovali sme nominálnu frázu:

krídl-
LOKATÍV» *juh-*
PARTITÍV» *hrad-*

Nejednoznačnosť, či táto nf je v nominatíve (subjekt) alebo akuzatíve (objekt), stále zostáva. Čisto syntaktický analyzátor by neskôr rozpoznal, že nemôže ísť o subjekt, pretože sa nezhoduje s centrálnym slovesom v čísle. Ale prvé dva riadky práve používaného slovesného intenčného rámca toto rozhodnú bezprostredne. Krídlo hradu je nepremiestňovateľné miesto. Z tejto charakteristiky nie je podľa hierarchie sémantických príznakov inferovateľná ani 'životnosť', ani 'inštrumentálium', ktoré sú podmienkou pre subjektové roly AGENSa alebo INŠTRUMENTu. Teda nf *južné krídlo hradu* je v role OBJEKTu.

Ďalej prejdeme hlavné sloveso *postavili* (hrana 2), zapamätajúc si jeho gramatické kategórie. Neskôr zaevidujeme ich zhodu s osobou a číslom subjektu *majstri*.

Ďalej. Predložková fráza (hrana 4) nenasleduje. Očakávame a nachádzame nf *Florentskí majstri*. Sufix *-ský* indikuje (pravidlo LVAL č.3), že ide o pôvod *majstrov* (premiestňovateľné osoby), nie ich lokáciu. Gramatická zhoda prídavného a podtatného mena sedí, fráza je v nominatíve, životná, teda dostane rolu AGENS (1. riadok valenčného rámca):

AGENS» *majster-*$^{+pl}$
ORIGINATÍV» *Florenci-*

Nasleduje predložková fráza *z guľatých skál*. Použitím 4. pravidla LVAL dostávame:

skal-$^{+pl}$
RELÁTOR» guľ-

Sufix *-atý* má špecifický význam: byť taký, ako niečo (guľatý = ako guľa). Táto pf by bola mohla rozširovať aj predchádzajúcu nf. Ale RVAL pravidlo č. 3 to nedovoľuje. *Florentskí majstri* nemôžu byť *z guľatých skál*. Pád MATERIÁL sa pripája len k neživotným substanciám.

Siedmy riadok valenčného rámca umožňuje podradiť túto pf pod centrálne sloveso so sémantickou rolou MATERIÁL. Konečná analýza vyzerá takto:

MANUÁLNY_KONŠTRUKTÍVNY_PROCES:

stav-$^{+pret\ +perf}$
AGENS» *majster-*$^{+pl}$
ORIGINATÍV» *Florenci-*
PACIENS» *krídl-*
LOKATÍV» *juh-*
PARTITÍV» *hrad-*
MATERIÁL» *skal-*$^{+pl}$
RELÁTOR» *guľ-*

6.6 Principiálne ohraničenia sémantických gramatík

Sila každej sémantickej gramatiky spočíva v tom, že stelesňuje určitý systém poznatkov o fyzikálnej povahe sveta. Vďaka tomu môže a musí zamietnuť každú frázu, ktorá týmto poznatkom odporuje:

kameň rozpráva
lietajúci zámok
kocúr kráľom

Takéto frázy sa však v jazyku používajú a človek s nimi bez problémov a zmysluplne narába. Majú význam, aj keď nemajú obsah. Vyskytujú sa pri opise rozprávkových alebo vedecko-hypotetických svetov s inými fyzikálnymi zákonmi.

Sémantická gramatika musí vylúčiť aj všetky metaforické významy. No metaforickosť v jazyku nie je žiaden marginálny, ale centrálny jav (Lakoff - Johnson, 1980). Celá štruktúra ľudského myslenia je prestúpená metaforami a analógiami.

Nechať sémantickú gramatiku akceptovať aj frázy s neštandardnou spájateľnosťou by znamenalo anulovať celú jej užitočnú hodnotu, pretože potom by akceptovala *všetky* frázy bez výnimky. Nechať gramatiku akceptovať iba frázy prípustné v nejakom fyzikálnom svete znamená urobiť ju strnulou, pretože nám potom zabraňuje myšlienkový experiment za hranice tohto sveta.

Život je fenomén na hranici chaosu a pevnej štruktúry a jazyk na tento nepretržitý pohyb živo reaguje neustálou asimiláciou a akomodáciou foriem aj významov na všetkých úrovniach. Každá konečná klasifikácia by bola ustrnutím, ktoré by od tej chvíle znamenalo odtrhnutie od života. Škatuľková sémantika principiálne nemôže urobiť zadosť nekonečnej variabilite a nepretržitému pohybu javov okolo nás. Množstvo vedomostí, ktoré sú stelesnené v pravidlách, bude vždy obmedzené. A aj keby sme skonštruovali nekonečne senzitívnu gramatiku s neobmedzeným množstvom poznatkov, nedokázali by sme vyriešiť *problém vymedzenia* (McCarthy - Hayes, 1969), t.j. nevedeli by sme, podľa akého pravidla vybrať z množstva pravidiel tie, ktoré sú vhodné v aktuálnej životnej situácii. Všetky pravidlá by pre počítač navždy zostali niečím cudzím, zvonku daným, niečím, čo si nesformuloval sám na základe vlastných skúseností a o čom nevie, za akých okolností a na základe čoho to bolo sformulované. Prejaví sa to tak, že bude používať síce správne pravidlá, ale na nesprávnom mieste a v nesprávny čas.

Vnímať jazyk pre človeka neznamená len pasívne preberať informáciu, ktorá prichádza akoby daná. Počúvanie je tvorivá aktivita, počas ktorej význam slov aktuálne a nanovo tvoríme, konštruujeme na pozadí svojej vnútornej skúsenosti. Tu nejde iba o prehľadávanie vopred daných alternatív. Práve preto je ľudská bytosť schopná porozumieť metaforám a preneseným významom, ktoré počuje trebárs aj po prvý raz v živote; preto sme schopní chápať básne a rozumieť symbolom. Naplňujeme ich vlastným prežitím, cítením, myslením, vnímaním a konaním, od ktorých je celý jazykový proces neoddeliteľný.

Počítaču chýba práve toto vnútorné bohatstvo, z ktorého by mohol tvoriť a ktorým by mohol napĺňať prázdne jazykové formy. *Počítač by musel byť schopný tvorivo konštruovať nové významy, a to na základe vlastnej skúsenosti; nie iba vykonávať exhaustívne prehľadávania medzi alternatívami preddefinovanými programátorom.*

7.

DERIVATOLOGICKÁ ANALÝZA

7.1 Potreba slovotvornej analýzy

Pod vplyvom výskumu, ktorého ťažisko sa dlhý čas odohrávalo v anglofónnych a frankofónnych krajinách, sa v jazyku nedocenila funkcia tvaroslovia a najmä slovotvorby. Pozornosť sa sústreďovala na syntax a derivatológia sa odsúvala ako okrajový jav.

Avšak žiaden dialógový či parafrázovací informačný systém sa nemôže zaobísť bez toho, aby slovotvorbe priznal miesto, ktoré jej patrí. Je jedným z troch pilierových podsystémov jazyka (spolu so skladbou a tvaroslovím). 85 percent slovenských slov sú deriváty alebo kompozitá; jadro slovnej zásoby pozostáva asi len z 15-tich percent neodvodených slov. V kapitole 1 sme dokonca poukázali na to, že skladanie a odvodzovanie je rovnako dobre súčasťou aj analytických jazykov, hoci v nich ide o odvodzovanie od supletívnych, inojazyčných koreňov.

Derivatológia sa zvykla posúvať do lexikológie. No lexikológia nedokáže chápať slová ako živé štruktúry, ale berie ich ako petrifikované celky. Vety

Vidím vodopád. *Peter je dobrý spevák.*

nestačí reprezentovať ako:

vidieť
KOG» *ja*
FEN» *vodopád*

byť
IFT» *Peter*
IFR» *spevák*
ATR» *dobrý*

Dôvod, prečo je takáto reprezentácia nedostatočná, sa nám vyjasní, ak si uvedomíme, že úlohou nášho systému je parafrázovanie, odpovedanie na otázky alebo návrat informácie z databanky. Všimnime si, že púhy abecedný slovník nijako nezachytáva informáciu, že napríklad medzi slovami *úsvit* a *svitanie* by existovala akákoľvek súvislosť. Ak systém nepoznáva súvislosť medzi slovom *vodopád* a slovami *voda, padať* alebo medzi slovami *spevák* a *spievať*, a nedokáže ich dynamicky skladať a rozkladať, musí v plnení požiadaviek používateľa bezpodmienečne zlyhať.

Ot: Vidíš padať vodu?
Od: Nie, iba vodopád.

Ot: Poznáš niekoho, kto dobre spieva?
Od: Nie. Viem, že Peter je spevák, ale nemám informáciu o nikom, kto by vedel spievať.

Ak má byť systém schopný odpovedať na takéto otázky, nemôže odvodené slová a zloženiny považovať za jednoduché lexikálne jednotky, ale musí ich rozbiť na elementárne komponenty a určiť sémantické vzťahy medzi nimi:

vidieť
KOG» *ja*
FEN» *miesto*
«LOK *padať*
AGS»*voda*

byť
IFT» *Peter*
IFR» *osoba*
«PDK *spievať*
ATR» *dobre*

Je teda nevyhnutné, aby parafrázovač narábal s derivatológiou ako so živým systémom - rovnako ako so syntaxou.

7.2 Radikálový rozklad

Vykonať slovotvornú sémantickú analýzu znamená rozložiť slovo na *ireducibilné radikály*, t. j. na ďalej nerozložiteľné plnovýznamové korene a komponenty významu, a určiť sémantické vzťahy (pády) medzi nimi.

Zložené a odvodené slová sa rozpadajú na jeden, dva, tri, niekedy až päť plnovýznamových koreňov plus niekoľko komponentov významu. Množina sémantických rôl, v ktorých korene vystupujú, je a musí byť totožná s uvedenou množinou 66-tich sémantických pádov; inak by usúvzťažnenie a parafrázovanie medzi syntaxou medzislovnou (vetotvorbou) a vnútroslovnou (slovotvorbou) ani nebolo možné:

zlatovláska

dievč-
«PAR vlas-$^{+pl}$
ATR» zlat-

kníhkupec

osoba
«AGS kúp-pred-
PAC» knih-$^{+pl}$

kameňolom

miest-
«LOK lám-
PAC» kameň-

zámorie

miest-
«LOK_5 mor-

strieborniak

minc-
MAT» striebr-

dračica

drak^{+fem}

loďkovať sa

pohybov- sa
INS» loď-$^{+min}$

kraľovať

byť-
IFR» kráľ-

skrášliť

rob-$^{+perf}$
ATR» krásn-

čítavať

čít-$^{+iter}$

rozosmiať sa

smia- sa$^{+inc\ +perf}$

prebehnúť

bež-$^{+trans\ +perf}$

7.3 Funkcie CVAL, DVAL

Slovné korene, prefixálne a sufixálne formanty podliehajú určitým pravidlám skladania do slov, podobne ako slová podliehajú pravidlám skladania do viet. Slovotvorba je vlastne vnútrosłovná syntax (hyposyntax).

Pravidlá spájateľnosti koreňov, ich onomaziologických kategórií a afixov sú zachytené v kolokačných funkciách CVAL (kompozičné vzorce) a DVAL (derivačné vzorce). Gramatika SAPFO obsahuje 11 kompozičných a 55 derivačných vzorcov.

Kompozičná kolokačná funkcia CVAL má dvanásť argumentov:

1) Slovný druh ľavého koreňa.
2) Sémantická črta ľavého koreňa.
3) Derivačný afix ľavého koreňa.
4) Ľavý koreň.

5) Slovný druh pravého koreňa.
6) Sémantická črta pravého koreňa.
7) Derivačný afix pravého koreňa.
8) Pravý koreň.
9) Slovný druh kompozita.
10) Sémantická črta kompozita.
11) Derivačný afix kompozita.
12) Kompozitný koreň.

Na základe týchto dvanástich argumentov je CVAL schopná rozhodnúť o sémantickej štruktúre kompozita. Výsledkom je kus sémantickej siete, ktorý túto štruktúru vyjadruje. Napríklad pravidlo č. 2 vyzerá takto:

```
CVAL(lk(sub,Se,_,K1),rk(sub,osoba,Sf,K2),kk(sub,osoba,'',K3),S)
   :- (Sf=ec;Sf=ač;Sf=ič;Sf=ík;Sf=ca;Sf=ár;Sf=ar;Sf=iar;Sf=''),
      koreň(K2,K4,_,_),
      koreň(K4,_,sl,I),
      VAL(I,_,_,živ,SP1),subj_pád(SP1),
      prefer_pád(I,_,_,Se,SP2),
      S = [ K3
            «SP1 K4
                  SP2» K1 ].
```

Tento kompozičný vzorec je zodpovedný za analýzu a syntézu všetkých zložených pomenovaní osôb, ktorých ľavý koreň je substantívny s nejakou sémantickou črtou *Se* a pravý koreň je deverbatívne osobné substantívum, odvodené od slovesa s intenčnou triedou *I* jedným z derivačných sufixov *-ec, -ač, -ič, -ík, -ca, -ák, -ar,*

-iar alebo prázdnym sufixom. Ide teda o slovotvorný typ *zvonolejec, delostrelec, zemepisec, horolezec, kníhkupec, svetobežník, kovorobotník, chlebodarca* a pod.

CVAL pritom volá funkciu VAL, aby rekonštruovala sémantický vzťah medzi slovesom v pravom koreni a substantívom v ľavom koreni práve tak, akoby substantívum bolo aktantom slovesa vo vete. Pretože predložkový pád pri volaní VAL nešpecifikujeme (po kondenzácii už v kompozite nie je viditeľný), VAL vráti tzv. prioritný pád, t. j. prvý nájdený pád v intenčnom rámci slovesa, ktorý unifikuje so sémantickou črtou aktanta. Napríklad vzťah medzi *strieľať* - *luk* identifikuje ako INŠTRUMENT, *liať*$_2$ - *zvon* REZULTANT, *liezť* - *hora* DESTINATÍV, *písať* - *zem* ako TEMATÍV. Nakoniec CVAL skonštruuje a vráti takéto štruktúry:

zvonolejec	*lukostrelec*	*horolezec*	*zemepisec*
osoba «AGS lia$_2$- REZ»zvon-	osoba «AGS strieľ- INS»luk-	osoba «AGS liez- DES»hor-	osoba «EDI pís- TEM»zem-

Ak sa naspäť k povrchovej štruktúre vrátime cez funkciu RVAL, dostávame parafrázy vo forme menných fráz:

zvonolejec	zemepisec
osoba, čo leje zvon	učenec, čo píše o zemi
lukostrelec	**horolezec**
osoba, čo strieľa z luku	osoba, čo lezie na horu
osoba, čo strieľa lukom	osoba, čo lezie po hore
	osoba, čo lezie do hory
	*osoba, čo lezie k hore

Derivačná kolokačná funkcia DVAL má 8 argumentov:

1) Slovný druh derivátu.
2) Sémantická črta derivátu.
3) Derivačný prefix.
4) Derivačný sufix.
5) Derivát.

6) Slovný druh koreňa.
7) Sémantická črta koreňa.
8) Koreň.

Napríklad pravidlo DVAL č. 1 má takýto tvar:

```
DVAL((sub,Se,Pr,Sf,K1),(sl,I,K2),S)
    :- (Sf=ec;Sf=ca;Sf=ič;Sf=ač;Sf=ák;Sf=iak;Sf=ér;Sf=ant;Sf=ník;
        Sf=čík;Sf=ár;Sf=iar),
       VAL(I,_,_,Se,SP),
       S = [osoba
            «SP K2]
```

Zodpovedá za deverbatívne substantívne mená osôb určitého druhu. Ide o typ *strelec, radca, tulák, vládca, kurič* a pod. Osoba vzhľadom na dej vystupuje v niektorej subjektovej role:

vládca	*radca*	*znalec*
[osoba «AGS vládn-]	[osoba «EDI rad-]	[osoba «KOG pozn-]

Pretože gramatika SAPFO je reverzibilná, pravidlá možno použiť zľava doprava, ako aj sprava doľava. Pri analýze sa používajú zľava doprava, t. j. premieňajú derivát na jeho hĺbkovú štruktúru. Túto hĺbkovú štruktúru možno znova premeniť na povrchové formy použitím tých istých pravidiel sprava doľava (syntéza). Štruktúra pravidiel ale umožňuje použiť ich pri syntéze viacerými spôsobmi:

strelec --analýza--> [osoba «AGS strieľ-] --syntéza--> *strelec, strelič, strelák, strelník, strelca, strelčík, strelár, strelér*

Samotné derivačné pravidlá pripúšťajú rovnako dobre aj *streliča* a *strelníka* ako pomenovanie subjektu deja. Treba uznať, že tieto novotvary sú naozaj derivatologicky v poriadku. Ale súčasťou kodifikovaného jazykového úzu je spomedzi všetkých možností iba slovo *strelec*. Vzniku neexistujúcich kompozít a derivátov zabraňuje slovotvorný slovník. V slovníku sa nachádzajú iba skutočne existujúce odvodeniny a moduly CVAL a DVAL svoje výsledky so slovníkom vždy konfrontujú.

7.4 Translexikálna syntaktická analýza

Bežná syntaktická analýza začína počiatočným symbolom pre vetu; rozkladá ju na stále menšie frázy a podfrázy a končí slovami ako terminálmi. Translexikálna (cezslovná) syntaktická analýza sa nezastavuje na úrovni slov, ale pokračuje až po ireducibilné slovné korene, slovotvorné formanty a elementárne komponenty významu, ktoré už nemožno ďalej rozložiť. Výsledkom je sémantická sieť, ktorá

pozostáva iba z najmenších samostatne vyjadriteľných významových komponentov. Počet aktantov v tejto sieti sa nerovná počtu slov vo vete, ale počtu plnovýznamových koreňov a afixov, ktoré plnovýznamové korene zastupujú. Syntakticko-slovotvorná analýza SAPFO je teda jednoliaty proces, pri ktorom sa začína medzislovnou a končí vnútroslovnou syntaxou, ktorá je jej živým predĺžením; obe sú spracúvané aj v rámci jednotného formalizmu (kolokačnými funkciami).

Medzi touto nad- a pod-slovnou syntaxou existujú určité rozdiely:

- Vnútroslovná syntaktická analýza postupuje sprava-doľava, nie zľava-doprava. Kompozitum a derivát sa podobajú zrkadlovému odrazu vety.
- Časti slova nie sú vyznačené medzerami, ale splývajú.
- Namiesto plnovýznamových slov môžu stáť afixálne formanty.
- Po kondenzácii do kompozita alebo derivátu sa časť informácie stratí.

7.5 Prelínanie supra- a sublexikálnej úrovne

Skúmajme syntagmy

dobrý vodič *dobrácky vodič*

Povedať, že adjektívum je determinatívnym prívlastkom, atribútom podstatného mena, je nedostatočné:

vodič-
ATR» *dobr-*

vodič-
ATR» *dobráck-*

Pretože v druhom prípade je to subjekt vedenia, ale v prvom samotný proces vedenia, ktorému sa pripisuje atribút:

vies-
AGS» *osob-*
ATR» *dobr-*

vies-
AGS» *osob-*
ATR» *dobráck-*

S ohľadom na to sa musia líšiť aj príslušné parafrázy:

Ján je dobrý vodič.
Ján jazdí dobre.
Jánova jazda je dobrá.

Ján je dobrácky vodič.
Ján je dobrácky.

a možné dialógy by boli:

Ot: Ako jazdí Ján?
Od: Dobre.
Ot: Aký je Ján?
Od: Neviem.

Ot: Aký je Ján?
Od: Dobrácky.
Ot: A ako jazdí?
Od: Neviem.

Takéto dialógy nie sú zvládnuteľné pre žiaden systém, ktorý by ignoroval, že na sublexikálnej úrovni sa vnútri slova *vodič* v skutočnosti nachádzajú dva participanty: subjekt vedenia a proces vedenia.

Zvláštne na tomto príklade je, že determinatívna supralexikálna syntagma „dobrý - osoba" sa v prvom prípade vnára až do predikatívnej sublexikálnej syntagmy „osoba - viesť". Formálne *dobrý* determinuje *Jána*, ktorý je hlavou sublexikálnej syntagmy *vod-ič*, súc v nej zastúpený substantívnym formantom *-ič*. Významovo však *dobrý* determinuje sloveso *viesť*, ktoré je syntakticky podradené až o jednu úroveň nižšie.

Slovom tieto dve syntagmy, slovotvorná a syntaktická, sa zlievajú a jednu nemožno analyzovať bez druhej. Trojicu participantov - *Ján, viesť, dobrý* - musíme uvažovať naraz, aby sme dokázali určiť ich vzájomné vzťahy. To však narúša náš doterajší postup, pri ktorom sme syntagmy analyzovali vždy iba ako bilaterálne dvojice, jednu po druhej.

7.6 Viacznačnosť a vágnosť slovotvorného významu

Viacznačnosť. Pri slovotvornom postupe nastáva v porovnaní so syntaxou určitá strata informácie, ktorá sťažuje jednoznačnú spätnú analýzu. Napríklad slovo *nafták* môže znamenať motor alebo auto, ktoré používa (spaľuje) naftu, alebo niečo, čo je vyrobené z nafty (ako *drevák*). *Vinár* môže byť niekto, kto víno buď dorába, alebo ho len pije a pod.

Zamlčané časti významu sa neraz podarí doplniť vďaka synergickému efektu: spoluprácou s asociačným, hyperonymickým, synonymickým a štylistickým slovníkom ako aj skúmaním intenčných štruktúr a spájateľnosti zúčastnených koreňov (pozri príklady 7.10 - 7.15).

Ak ani celková informácia v systéme na jednoznačnú rekonštrukciu slovotvorného významu nedostačuje, slovotvorná analýza zostáva viacznačná a taká sa odovzdáva ďalej. Konečné rozhodnutie sa prenecháva kontextovej analýze alebo inferencii.

Vágnosť. Iným prípadom je vágnosť slovotvorného významu. Napríklad formant *-ník* je prípona s veľmi širokým, neurčitým významom:

slnečník; denník; čajník; kávovník; makovník; skleník
skladník; pilník; rybník; vodník; rovník; poludník
zásobník; nárazník; stotník; večerník; veterník; dvojník

Zatiaľ čo viacznačnosť sa prejaví viacerými výslednými analýzami, vágny význam sa prejaví vo výslednej analýze prítomnosťou vágneho sémantického pádu. V našom systéme sa takto používajú dva pády: OKUPATÍV (zaoberať sa niečím) a KOHERENTÍV (neurčitá súvislosť). Napríklad motivácia slov *rovník, poludník* je taká nezvyčajná, že ju nemožno vyjadriť žiadnym zo špecifických sémantických

pádov v našom systéme. Vzťah medzi *poludním* a *poludníkom* sa teda reprezentuje pomocou sémantického pádu KOHERENTÍV. Takáto analýza potom nie pre ďalšie spresnenie uzavretá. Neurčitý pád nemusí, ale môže byť neskôr s ohľadom na novšie informácie nahradený pádom špecifickejším.

Obojznačnosť. Vo zvláštnych prípadoch môže byť výsledok slovotvornej analýzy dvoj- alebo viacznačný, pričom obe (všetky) alternatívy sú platné súčasne. Napríklad sufixy *-ár, -eň* alebo *-ár, -stvo* splynuli do nového slovotvorného sufixu *-áreň, -árstvo* s miestym významom. Podľa toho *voňavkárstvo* možno chápať ako *miesto, kde sú voňavkári*, ale aj ako *miesto, kde sú voňavky*:

voňavkárstvo

miest-
«LOK_1 = OKU
«LOK_1 osob-$^{+pl}$
OKU» látk-$^{+pl}$
«PDK voň-$^{+incl}$

V iných prípadoch by sa žiadal obojznačný sémantický pád. Napríklad *zväčšovák* možno chápať ako subjekt aj ako nástroj deja (*zväčšovák zväčšuje, zväčšovákom sa zväčšuje;* riešenie je v kapitole 12).

7.7 Petrifikácia, lexikalizácia a idiomatizácia

Slovná zásoba podlieha vývoju a metaforickým posunom. Zložené slovo môže neskôr znamenať niečo viac, niečo menej alebo niečo iné ako pôvodné zložky, z ktorých sa skladá. Lexikalizovaný význam nemusí viac zodpovedať významu, ktorý by naznačovala slovotvorná forma. Napríklad *zámočník* nepracuje iba so zámkami, ale so železiarskymi výrobkami všeobecne.

Niektoré lexikalizácie sa darí identifikovať a korigovať v súčinnosti s ostatnými jazykovými podsystémami (pozri príklady 7.13, 7.14 a 7.15).

Vo zvyšných prípadoch zostane slovotvorná analýza čiastočne nesprávna. Napríklad slovo *zelenina* dostane rovnakú výslednú analýzu ako *zelina:* ako kolektívum pre *zelené rastliny*. Avšak v prípade zeleniny nie každá zelená rastlina je zelenina, ani nie každá zelenina je zelená.

Nedostatky analýzy sa potom zreteľne prejavia pri parafrázovaní:

drahý náhrobný kameň	
náhrobný drahý kameň	/Fok
drahý kameň na hrobe	/Nor
drahokam na hrobe	/Kpz

7.8 Slovotvorná parafráza na počítači

Pri tvorbe počítačového modelu slovenskej slovotvorby sme zužitkovali materiál z prác J. Horeckého (1959, 1954, 1971) a z Mluvnice češtiny (1986). Obsahujú klasifikovaný materiál a príklady hotových parafráz. Neobsahujú však návod, *ako* by mohol slovenské kompozitá a deriváty klasifikovať a parafrázovať niekto, kto to nemôže urobiť *bez rozmýšľania*, pretože sa nenarodil ako Slovák. Ak chceme, aby počítač samostatne tvoril parafrázy, musíme podať úplné a explicitne formulované pravidlá, zákonitosti, ktorým je slovotvorné parafrázovanie podriadené. Nič nemožno ponechať bez vysvetlenia ako „intuitívne zrejmé".

Napríklad *moreplavba, jazykoveda, olejomaľba* znamenajú *plavbu* ***po*** *mori, vedu* ***o*** *jazyku, maľbu olejom. Zemepisec, horolezec, delostrelec, meďorytec* sa parafrázujú ako ten, kto *píše* ***o*** *zemi, lezie* ***po*** *horách, strieľa* ***z*** *dela, ryje* ***do*** *medi.* Tieto rozdiely nestačí vysvetliť vždy len *ex post*, od prípadu k prípadu. Treba vyznačiť *všeobecne platné pravidlo*, ktoré umožní chýbajúcu predložku *vopred* vypočítať. Zo slovies *hostiť/hosťovať*, samozrejme, prvé patrí medzi predmetové (spolu s *mrzačiť, väzniť, kresťančiť*) a druhé medzi podmetové (spolu s *kraľovať, záhradníčiť*). Ale podľa akého formálneho príznaku by sa tento rozdiel dal identifikovať, ak by sme si ho jednoducho ne*pamätali*?

Ako odlíšiť *mešťan* od *meštiak* ak sú korene totožné a slovotvorné sufixy *-an, -iak* synonymné?

Celý rad pokusov o navrhnutie formalizmu pre slovotvorné parafrázovanie zostal neúplný. Samotní autori sú si vedomí, že ich modely sú funkčné iba sčasti. Chceme ukázať, že jedným z limitujúcich faktorov bola hlavne snaha vypracovať takýto formálny aparát v rámci nejakej ohraničenej, jednoliatej podoblasti a obísť nevyhnutnosť zahrnúť do formálneho modelu jazyk ako celok.

Podľa W. Doroszewského (1962) slovotvorná štruktúra zodpovedá syntaktickým vetným členom. M. Dokulil (1962) upozorňuje, že to tak je iba niekedy a navrhuje vychádzať skôr z onomaziologických kategórií. Prečo by sa však riešenie malo nachádzať v rámci nejakej jednorozmernej teórie? Musíme pripustiť, že slovotvorná parafráza závisí od súhry celého radu heterogénnych faktorov súčasne. V našom modeli ich preto musíme aj takto uvažovať. J. Furdík (1978) správne navrhuje zblížiť oba prístupy tak, aby sa syntaktické štruktúry nepovažovali za východisko slovotvorných, ale za paralelné či komplementárne k slovotvorným, a hľadala sa tretia, hĺbková štruktúra, z ktorej obe tieto štruktúry vychádzajú.

O. G. Revzinová (1969) navrhuje robiť perifrastické substitúcie podľa vzorcov: *ten, kto..., to, že..., ten, čo...* (*strelec → ten, kto strieľa, streľba → to, že sa strieľa, trestanec → ten, čo je trestaný*). To je však formulované iba so zreteľom na deverbatívne a deadjektívne substantíva a aj tam je toto jednoduché delenie úplne nedostatočné. Popis sémantickej štruktúry deverbatív, ktorý by mohol slúžiť ako základ pre tvorbu parafráz, vypracovala K. Buzássyová (1974). Pri desubstantívach sa však rovnaký formálny aparát nedá použiť. Práve pri desubstantívach je zreteľné, že ich nemožno riešiť bez toho, aby sme uvažovavali súčinnosť širších súvislostí. W. Doroszewski (1962), O. G. Revzinová (1969), M. Dokulil (1962) a celý rad ďal-

ších autorov (Brodowska-Honowska, 1967) na klasifikáciu desubstantív buď rezignujú, alebo ju spracúvajú neucelene.

J. Furdík (1978) predkladá 22 parafrázovacích vzorcov. Zo všetkých slovných druhov však spomína iba odvodené substantíva a pri desubstantívnych substantívach konštatuje, že „medzi základom a formantom slova môžu byť najrozličnejšie vzťahy, ktoré sa z derivátu 'vyčítať' nedajú." Prečo ich však nevyčítať odinakiaľ ako z derivátu samotného? V jazyku predsa nič neexistuje osamote, ale vždy iba vo vzťahoch!

7.9 Synergický efekt pri slovotvornom parafrázovaní

Z rovnakého dôvodu sa Slovenská lexikológia (Horecký, 1971) a Mluvnice češtiny (1986) pri odvodených a zložených adjektívach obmedzili len na hrubú klasifikáciu a konštatovanie, že „vztahová adjektiva sama o sobě označují pouze sémanticky nediferencovaný vztah". Prečo by sa však práve adjektíva mali skúmať „samy osebe"? Samy osebe sa ani nevyskytujú. Už samo označenie „vzťahové", „prídavný" svedčí o tom, že ich význam vyplýva viac zo vzťahov k iným slovám, ako z nich samých. Napokon *všetky* slovné druhy majú širší význam, ktorý si navzájom zužujú, až keď stoja vedľa seba v kontexte.

Pretože konštatovanie, že význam adjektív je neurčitý, neprináša nič pozitívne, pracujeme radšej s množinou *potenciálnych, ale* ***určitých*** adjektívnych významov, ktoré sa aktualizujú v prítomnosti substantíva. Potenciálne sémy adjektív a substantív navzájom interagujú a vzájomne sa selektujú podľa určitých zákonitostí, ktoré možno skúmať a do určitej miery formalizovať. Prehľad potenciálnych významov adjektív vo vzťahu k substantívam podáva M. Nábělková (1989). Detailný, hoci tiež ohraničený popis interakcie adjektívnych a substantívnych významov nájdeme v poľskej derivatológii (Brodowska-Honowska, 1967).

Vzájomná prepojenosť sa prejavuje na všetkých úrovniach jazyka. Všimnime si, že morfologická analýza nie je možná bez syntaktickej. O slove *oko* nemožno povedať ani v akom je páde, ani aký má význam, ak nestojí minimálne v takom kontexte ako *...oko a ucho sú...* Pri slove *vila* neoznačíme ani slovný druh.

Podobne aj syntax nie je mysliteľná inak, ako sémantická syntax. Ani jedno pravidlo pre syntaktické transformácie, transpozície, konverzie alebo nominalizácie by sa nedalo formulovať, keby sme zároveň nevymedzili sémantické triedy, na ktorých funguje. Ani len syntaktická štruktúra vety by sa nedala identifikovať mimo lexikálnej sémantiky.

Slovotvorná analýza taktiež nie je oddeliteľná od syntaxe a sémantiky. V príklade 6.5 sme analyzovali význam adjektíva *južný* vo fráze *južné krídlo hradu.* Lokatívny význam slova *južný* závisel od príznaku premiestňovateľnosti slova *krídlo*, ten však, kvôli polysémickosti *krídla,* závisel od sémantiky slova *hrad.*

Nakoniec, morfológia, syntax, sémantika a slovotvorba napospol nie sú oddeliteľné od lexikálno-paradigmatického systému, od frazeológie, štylistiky, pragmati-

ky a všetkých ostatných jazykových podsystémov. Existencia nejakého synonyma, antonyma alebo frazeologizmu môže zasiahnuť dajme tomu do syntaxe a zmeniť celú konštitúciu vety.

Tvrdíme teda, že ani slovotvornú parafrázu, ani iné javy nemožno modelovať oddelene od všetkých ostatných podsystémov jazyka, ale iba vcelku. Vzniká tým *synergický efekt,* t. j. schopnosť systému vykonať viac, ako je suma toho, čo by jednotlivé časti systému dokázali vykonať oddelene.

Nasledujúce príklady strojovej analýzy slovotvorby zviditeľňujú dynamickú prepojenosť všetkých jazykových podsystémov, ktoré v súčinnosti tvoria jeden živý celok.

Žiadne statické, jednoduché vzorce slovotvornú parafrázu nezvládnu, a to bez ohľadu na ich počet. Program SAPFO obsahuje v súčasnosti 66 dynamických pravidel slovotvornej analýzy. Každé z nich generuje desiatky rozličných povrchových foriem v závislosti od momentálneho stavu ostatných zložiek a funkcií systému (vrátane stále sa meniacej bázy poznatkov), s ktorými nepretržite komunikuje.

Všimnime si, od akého počtu rozličných faktorov závisí správne utvorenie parafrázy. Rozhodujúce faktory sa líšia od prípadu k prípadu. Len kategória slovného druhu svojimi kombináciami vytvára vyše tridsať hrubých slovotvorných typov. K tomu sa pridružuje význam desiatok prefixov, sufixov a cirkumfixov. Význam ďalej ovplyvňujú sémantické triedy zúčastnených zložiek. To ale nie je všetko. Na parafrázu nestačia iba informácie o fundujúcom a fundovanom slove! Treba uvažovať aj onomaziologické kategórie a väzobné vlastnosti slov, ktoré v slovotvornom hniezde genealogicky predchádzajú alebo nasledujú, alebo sa v ňom nachádzajú na inom mieste. Parafrázovač musí úzko spolupracovať so slovníkmi: synonymickým, štylistickým, frazeologickým a so slovníkom lexikálnych parametrov (Apresian, 1974), ktoré poskytujú množstvo informácie o postavení slova a jeho vzťahoch v rámci celej lexiky. Nakoniec, do vzhľadu parafrázy môže zasiahnuť aj inferencia s bázou poznatkov. Vzájomná prepojenosť kolokačných funkcií SAPFO (VAL, LVAL, RVAL, CVAL, DVAL, AVAL) je taká jemná, že ak sa zmení kontext alebo pribudne nejaký nový poznatok v konotačných slovníkoch, slovotvorná parafráza môže už vyzerať úplne inak.

Náš počítačový model pokladáme za tvorivé nadviazanie na záverečné poznámky J. Furdíka (1978), na práce J. Horeckého (1978) a K. Buzássyovej (1974) pomocou výkonného nástroja - výpočtovej techniky.

SAPFO je rozsiahly systém programov (1 MB v jazyku Prolog). Ukázať úplný príklad všetkých procesov, ktoré sa aktivujú pri spracovaní vety, nie je možné. V príkladoch však uvedieme relevantné časti z valenčného, slovotvorného, lexikálno-sémantického, konotačného, frazeologiockého a štylistického slovníka, ktoré analýzu ovplyvňujú. Uvedieme aj všetky kolokačné pravidlá, ktoré sa pri analýze toho-ktorého príkladu aktivujú. Nebudeme popisovať prácu syntaktického a morfologického modulu, ale sústredíme sa na kolokačné funkcie, ktoré majú v celom procese riadiacu úlohu.

7.10 Kontrastívny príklad

zlatka / kráska

Pri konfigurácii:

zlato	(mat)	krásny	(kval)
zlatý$_1$	(klas)	kráska	(osoba)
zlatka	(vec)	prekrásny	(kval)
zlatonosný	(akvzt)	čarokrásny	(kval)

zlatý$_2$	(kval)	synon(zlatý$_2$,vzácny)
zlatučký	(kval)	hyperon(zlatka,minca)

```
DVAL((sub,Se,_,Sf,K1),(adj,_,K2),SP)
      :-(Sf=ec;Sf=ník;Sf=ík;Sf=ák;Sf=iak;Sf=ak;Sf=ka;Sf=ok;
         Sf=áň;Sf=ena;Sf=ina;Sf=uša;Sf=uľa),
        hyperon(K1,K3),
        LVAL(K2,_,K3,_,SP).

LVAL((adj,klas,_,Sf),(sub,mat),(sub,Se,_,_),_,mat)
      :-(Sf='';Sf=ov;Sf=en;Sf=n),sema(Se,kon),sema(Se,neživ).

LVAL((adj,kval,_,_),_,(sub,Se,_,_),_,atr).

RVAL((sub,Se),_,0,gen,(sub,atb),(adj,kval),atr).

RVAL((sub,Se1),_,z,gen,(sub,Se2),_,mat)
      :-sema(Se1,kon),sema(Se2,mat).
```

Dostávame parafrázy:

zlatka		kráska
zlatá minca	zlatá žena	krásna žena
minca zo zlata	vzácna žena	žena veľkej krásy

Uvedené pravidlo DVAL stelesňuje triedu deadjektívnych substantív, utvorených derivačnými sufixmi *-ec*, *-iak* atď. *Zlatka* aj *kráska* budú spracované obe týmto pravidlom, pretože patria do tej istej slovotvornej triedy.

Toto pravidlo pracuje tak, že vyhľadá hyperonymum k fundovanému substantívu. Pre *zlatka* sa v taxonomickej hierarchii nájde hyperonymum *minca*. Pre *krásku*

sa však žiadne hyperonymum v báze údajov nenachádza. Program teda utvorí hyperonymum podľa najabstraktnejších kategórií: sémantického príznaku (osoba) a rodu (žena). Derivačná funkcia DVAL takto rozvinula odvodeniny *zlatka/kráska* na menné frázy *zlatá minca/krásna žena*. Ľavostranné prívlastky však už patria do kompetencie nominalizačnej funkcie LVAL, ktorá je na ne špecialistkou. DVAL teda požiada o riešenie LVAL a odovzdá jej príslušné parametre. Tým sa úloha funkcie DVAL končí.

Až tu sa osudy *zlatky* a *krásky* začínajú rozchádzať. *Zlatá minca* unifikuje s prvým ľavovalenčným pravidlom. Ide o klasifikačné adjektívum, odvodené od materiália prázdnym sufixom. *Minca* vyhovuje podmienkam konkrétnosti aj neživotnosti. *Krásna žena* neprejde prvým pravidlom, pretože fráza je životná, ale prejde druhým, pretože ide o kvalitatívne adjektívum.

Zlatá minca dostane pád MATERIÁL, *krásna žena* ATRIBÚT.

Ak teraz tieto hĺbkové pády vyjadríme použitím pravidiel RVAL sprava doľava, t. j. budeme vychádzať od významu a pozrieme sa na formu, dostaneme parafrázy *minca zo zlata*, *žena veľkej krásy*. Od *zlatá žena* však dostaneme *vzácna žena*, nie *žena zo zlata*.

7.11 Kontrastívny príklad

ľudovláda / svetovláda / krutovláda / krvotvorba

Lexikálno-sémantický a slovotvorný slovník:

vládnuť	(147)	človek	(osoba)
vláda	(dej)	ľud	(osoba)
svetovláda	(dej)	ľudovláda	(dej)
ľudovláda	(dej)		
krutovláda	(dej)	svet	(lok)
		svetovláda	(dej)
tvoriť (20)		svetovládny	(akvzt)
tvorba	(dej)		
krvotvorba	(dej)	krv	(mat)
		krvný	(klas)
krutý	(kval)	krvavý	(akvzt)
kruto	(mds)	krvotvorba	(dej)
krutosť	(atb)		

Valenčný slovník:

VAL(147,0,nom,živ,ags).
VAL(147,0,dat,_,pac).

VAL(20,0,nom,_,ags).
VAL(20,0,ak,_,rez).

```
VAL(147,nad,inš,_,pac).
VAL(147,prísl,_,mod,mds).

CVAL(lk(sub,Se,_,K1),rk(sub,dej,Sf,K2),kk(sub,dej,'',K3),SP)
    :- (Sf='';Sf=a;Sf=ba;Sf=oba;Sf=ota),
       koreň(K2,K4,_,_),
       koreň(K4,_,sl,I),
       prefer_pád(I,_,_,Se,SP).

CVAL(lk(adj,_,_,K1),rk(sub,_,_,K2),kk(sub,_,'',K3),SP)
    :- LVAL(K1,_,K2,_,SP).

LVAL((adj,kval,_,_),_,(sub,dej,_,_),_,mds).

RVAL((sub,dej,_),(sl,I,_),0,gen,(sub,Se,_),_,SP)
    :- prefer_pád(I,_,_,Se,SP).
```

Výsledné parafrázy:

ľudovláda	svetovláda	krutovláda	krvotvorba
vláda ľudu	vláda svetu	krutá vláda	tvorba krvi
ľud vládne	vládnuť nad svetom	vládnuť kruto	tvoriť krv

V kompetencii prvého pravidla je analýza všetkých kompozít zložených z dvoch substantív, pričom druhé je dejové a výsledný kompozitný koreň je tiež dejové substantívum. Navyše pravý fundujúci koreň musí byť derivát utvorený slovotvorným sufixom *-ba, -oba, -ota* od slovesa. V intencii tohto slovesa potom funkcia CVAL hľadá preferovaný pád.

V intencii slovesa *tvoriť* dominuje akuzatív priameho objektu. Podľa toho *krv* je REZULTANT *tvorenia*.

Pri slovese *vládnuť* sa akuzatív nenachádza. Najdominantnejším je nominatív subjektu, preto *ľud* bude AGENS *vládnutia*.

V prípade *svetovláda* sa pokus umiestniť *svet* do subjektovej roly nepodarí, pretože valenčný riadok pre AGENS vyžaduje životnosť. Najdominantnejšia vyhovujúca rola je nepriamy objekt, vyjadrený holým datívom alebo predložkovým pádom *nad* + inštrumentál. *Svet* je PACIENS *vládnutia*.

Kompozitum *krutovláda* neunifikuje s prvým pravidlom, ale s druhým, pretože jeho ľavý koreň je adjektívny. Funkcia CVAL poprosí LVAL o riešenie prípadu *krutá vláda*. LVAL identifikuje ich vzťah ako MODUS. Vyjadruje sa spôsobovou príslovkou.

7.12 Kontrastívny príklad

hostiť / hosťovať

```
hosť            (osoba)
  hosťovský     (akvzt)
  hosťovať      (400)
  hostiť        (401)
    pohostiť    (401)
```

```
DVAL((sl,I,'',Sf,K1),(sub,Se,K2),SP)
   :-  (Sf=iť;Sf=ovať),(I=400;I=401),
       if koreň(K2,K3,sl,(sl,I,Pf,'',_)),
          (Pf=po;Pf=z;Pf=s;Pf=u;Pf=do),SP=pac);
       else
          SP=ags.
```

hosťovať	hostiť
byť hosťom	robiť hosťom

Slovies *hostiť/hosťovať* sa ujme derivatologické pravidlo pre predmetové a podmetové slovesá. Toto pravidlo stelesňuje poznatok, že od predmetových slovies sa ďalej prefixáciou odvodzujú dokonavé slovesá (*zmrzačiť, dokaličiť, pokresťančiť, uväzniť*). Od podmetových (*kraľovať, chalupárčiť*) sa neodvodzujú a ak áno, tak pomocou iného inventára predpôn.

7.13 Kontrastívny príklad

mešťan / meštiak

```
mesto           (lok)
  predmestie    (lok)
  mestský       (akvzt)          synon(meštiacky,snobský)
  mešťan        (osoba)          štyl(malomeštiak,pejor)
    mešťanský   (klas)
  meštiak       (osoba)
    malomeštiak (osoba)
    meštiacky   (akvzt)
```

```
DVAL((sub,osoba,Pf,Sf,K1),(sub,lok,_),SP)
   :-  (Sf=an;Sf=čan;Sf=ák;Sf=ak;Sf=iak;Sf=ár;Sf=ar;Sf=iar),
       (Pf=zá;Pf=prí;Pf=pod),
   if (koreň(K1,K2,adj,akvzt),synon(K2,K3),koreň(K3,_,adj,kval))
   or (koreň(K1,K2,_,_),štyl(K2,_)),
       SP=rel,
   else
       SP=ori.
```

mešťan	meštiak
človek z mesta	človek ako z mesta
	snobský človek

Hoci slovotvorné sufixy *-an, -iak* sú synonymné, *meštiak* má oproti *mešťan* pejoratívny kvalitatívny odtienok. DVAL identifikuje tento rozdiel podľa toho, že *meštiak* má v slovotvornom hniezde uvedené stupňovateľné adjektívum *meštiacky*, ktoré stojí vo vzťahu synonymie s čisto kvalitatívnym adjektívom; a tiež slovo *malomeštiak*, ktoré je podľa informácie zo štylistického slovníka štylisticky príznakové. Slovotvorná parafráza sa teda v tomto prípade nevykonáva na základe slov samotných, ale na základe ich konotácií.

7.14 Kontrastívny príklad

telefonovať / bicyklovať sa		*krúžiť / krúžkovať*	
telefón	(ins)	kruh	(lok)
telefonovať	(60)	krúžiť	(1)
telefonicky	(mod)		
		kruh	(vec)
bicykel	(ins)	krúžok	(vec)
bicyklovať sa	(1)	krúžkovať	(444)

```
hyperon(telefonovať,hovoriť).
hyperon(bicyklovať sa,jazdiť).
hyperon(krúžiť,pohybovať sa).
hyperon(krúžkovať,opatriť).

VAL(60,0,nom,živ,edi).
VAL(60,0,dat,živ,ads).
VAL(60,0,ak,_,frm).
VAL(60,0,ins,ins,ins).
VAL(60,o,lok,_,tem).
VAL(60,cez,ak,lok,via).
VAL(60,cez,ak,ins,ins).
VAL(60,prísl,_,_,ins).
```

```
VAL(444,0,nom,živ,ags).
VAL(444,0,ak,kon,pac).
VAL(444,0,ins,vec,rea).

VAL(1,0,nom,_,ags).
VAL(1,do,gen,lok,des).
VAL(1,0,ins,lok,via).
VAL(1,0,ins,abs,mds).
VAL(1,0,ins,ins,ins).
VAL(1,0,na,lok,ins,ins).
VAL(1,v,lok,lok,lok).
VAL(1,po,lok,lok,lok).
```

```
DVAL((sl,I,'',ovať,K1),(sub,Se,_),SP)
    :- hyperon(K1,K2),
       špecif_pád(K2,_,_,Se,SP).

špecif_pád(I,Pr,Pa,Se1,SP)
    :- VAL(I,Pr,Pa,Se2,SP),min_taxon_vzdial(Se1,Se2).
```

krúžkovať	telefonovať
opatriť krúžkom	hovoriť cez telefón
	hovoriť telefónom
krúžiť	hovoriť telefonicky
pohybovať sa do kruhu	
pohybovať sa v kruhu	bicyklovať sa
pohybovať sa po kruhu	jazdiť bicyklom
*pohybovať sa ku kruhu	jazdiť na bicykli
*pohybovať sa od kruhu	*jazdiť v bicykli
*pohybovať sa na kruh	

Toto zdanlivo jednoduché pravidlo pre desubstantívne slovesá vďaka svojej dynamickej interakcii s ostatnými subsystémami generuje celú škálu rôznorodých parafráz. Postupuje tak, že v intenčnom rámci hyperonyma vyhľadá pád, ktorý je špecifický tým, že si vyžaduje sémantickú črtu substantíva, od ktorého je hyponymné sloveso odvodené. Tento algoritmus, pravda, zďaleka nepracuje bezúhonne. Napríklad ku slovesu *pohybovať sa* pripojil všetky miestne pády, hoci iba prvé tri sú prijateľné.

7.15 Kontrastívny príklad

vinár / stolár / organista / chemik

víno	(mat)	asoc(víno,piť)	
vinár	(osoba)	asoc(víno,dorobiť)	
		asoc(organ,hrať3)	
stôl	([kon,artef])	asoc(X,robiť) :- sema(X,artef).	
stolár	(osoba)	hyperon(chemik,vedec)	
		synekdocha(stôl,nábytok)	
organ	(ins)		
organista	(osoba)	robiť	(20)
		dorábať	(20)
chémia	(abs)		
chemik	(osoba)	piť	(500)
hrať3	(251)	zaoberať sa	(153)

```
DVAL((sub,osoba,_,Sf,K1),(sub,Se,K2),L)
    :-  (Sf=ník;Sf=čík;Sf=ík;Sf=ca;Sf=ár;Sf=ec;Sf=ér;Sf=ista),
        (asoc(K2,K3);K3='zaoberať sa'),!,
        (synekdocha(K2,K4);K4=K2),!,
        (hyperon(K1,K5);K5=osoba),
        koreň(K3,_,sl,I),
```

VAL(I,_,_,živ,SP1),subj_pád(SP1),
VAL(I,_,_,Se,SP2),obj_pád(SP2),

$$L = \begin{bmatrix} K5 & \\ & SP1: K3 \\ & & SP2: K4 \end{bmatrix}.$$

organista	vinár
osoba, ktorá hrá na organe	osoba, ktorá pije víno osoba, ktorá dorába víno
stolár	**chemik**
osoba, ktorá robí nábytok	vedec, ktorý sa zaoberá chémiou

Toto derivatologické pravidlo využíva asociačné štruktúry (1.9). Podľa návrhu J. Pustejovského (1991) priraďujeme substantívam systém slovies, činností, na ktorých typicky participujú. Takýchto typických činností môže existovať viac, alebo ani jedna. Napr. *ihla → šiť; kniha → čítať, písať; človek → 0.* Pri analýze desubstanív sa asociácie vynikajúco uplatnia, pretože vzťah medzi fundujúcim slovom a sufixom nepoukazuje na hociakú, ale práve na typickú činnosť.

Pri slove *stôl* konotačný slovník žiadnu asociáciu neposkytuje, ale nahradí ju všeobecné pravidlo, ktoré hovorí, že artefakty sú vždy objektom slovesa *robiť*.

Chémia nevyhovuje ani tomuto pravidlu, a tak DVAL supluje centrálne sloveso najvšeobecnejším možným: *zaoberať sa.*

Po rekonštrukcii centrálneho slovesa sa snaží interpretovať osobu ako subjekt deja a fundujúci koreň ako jeho objekt, resp. okolnosť.

Pravidlo berie ďalej do úvahy, či fundujúci koreň nefunguje ako synekdocha. *Stolár, zámočník* sa nezoberajú iba stolmi a zámkami, lebo *stôl, zámok* sú iba synekdochami drevených a železiarskych výrobkov.

Nakoniec DVAL vráti ako výsledok celú štruktúru - kus sémantickej siete. V našom príklade sú to tieto:

$$\begin{bmatrix} \text{osoba} & \\ \text{AGS »hrať} & \\ & \text{INS »organ} \end{bmatrix} \begin{bmatrix} \text{osoba} & \\ \text{AGS »piť} & \\ & \text{PAC »víno} \end{bmatrix} \begin{bmatrix} \text{osoba} & \\ \text{AGS »dorobiť} & \\ & \text{REZ »víno} \end{bmatrix}$$

$$\begin{bmatrix} \text{osoba} & \\ \text{AGS» robiť} & \\ & \text{REZ» nábytok} \end{bmatrix} \begin{bmatrix} \text{vedec} & \\ \text{AGS» zaoberať sa} & \\ & \text{PAC» chémia} \end{bmatrix}$$

7.16 Protipríklad

Slová *vodár, vodák, vodník* v sémantickom univerze SAPFO splývajú a dostanú rovnakú reprezentáciu (osoba, ktorá sa zaoberá vodou):

osob-
OKU» vod-

7.17 Štruktúra kolokačných funkcií

„Slovotvorný typ je daný jednotou formantu, onomaziologické struktury a lexikálně-gramatické kategorie základu" (Mluvnice češtiny, 1986). Jednotlivé onomaziologické kategórie a formanty sa vyberajú z určitých inventárov, ktoré možno vymenovať:

Prefixálny formant	Slovný druh	Séma	Sufixálny formant	Výsledný sl. druh	Séma	Sémantický pád
pod-	sub	osoba	-ec	sub	osoba	ags
pred-	adj	ins	-ič	adj	ins	pac
zá-	adv	lok	-ník	adv	lok	ins
sú-	sl	dej	-ie	sl	dej	lok
0	čís	mat	-osť	čís	mat	ori
vy-	zám	atb	-ba	zám	atb	atr
od-	část	korp	-iť	část	korp	mat
pra-	cit	flor	-ovať	cit	flor	tmp

Zadefinovať slovotvorný vzor s matematickou presnosťou, akú vyžaduje počítač, znamená pospájať prvky týchto siedmich inventárov onomaziologických príznakov a formantov všetkými prípustnými spojnicami. Treba taxatívne vymenovať všetky sedmice, ktorých n-tý člen je vybratý z prvkov n-tého inventára a ktoré sa v jazyku uplatňujú. Napr.:

sú	+	sub	+	kon	+	ie	+	sub	+	kon	=	kar	[typ súhvezdie]
0	+	sl	+	120	+	ec	+	sub	+	osoba	=	ags	[typ strelec]
0	+	adj	+	kval	+	ec	+	sub	+	živ	=	atr	[typ dravec]
0	+	sub	+	osoba	+	ovať	+	sl	+	400	=	ifr	[typ kraľovať]
u	+	sl	+	ndk	+	0	+	sl	+	dok	=	vid	[typ ušiť]

Derivačná kolokačná funkcia DVAL je teda v podstate sedmicou zreťazených onomaziologických kategórií.

Uvedeným spôsobom postupoval J. Horecký (1964), ale uvažoval iba formálne príznaky a nepospájal ich so sémantickými. Buzássyová (1974) takto spracovala deverbatíva.

Tento postup je základom všetkých šiestich kolokačných funkcií systému SAPFO:

VAL (valencia, väzobné vlastnosti slovies)
LVAL (ľavostranná spájateľnosť substantív, ľavé nominalizačné vzorce)
RVAL (pravostranná spájateľnosť substantív, pravé nominalizačné vzorce)
CVAL (spájateľnosť slovných koreňov a afixov, kompozičné vzorce)
DVAL (spájateľnosť slovných koreňov a afixov, derivačné vzorce)
AVAL (spájateľnosť adjektív)

Základom VAL je pätica onomaziologických kategórií, základom CVAL desatica, DVAL sedmica, AVAL dvojica. LVAL je tvorená trinásticou, RVAL jedenásticou.

Z príkladov vidíme, že kolokačné funkcie sú v skutočnosti oveľa komplexnejšie a vykonávajú celý rad špecifických výpočtov v súčinnosti so slovníkmi: konotačným, slovotvorným, lexikálno-sémantickým, synonymickým, frazeologickým a štylistickým. Ich jadrom je však vždy **n-tica zreťazených formálných a významových príznakov.**

Šestica kolokačných funkcií nám umožní spracúvať všetky tri podsystémy jazyka v rámci jediného formalizmu a tak modelovať ich vzájomnú prepojenosť.

VAL, LVAL, RVAL, AVAL zabezpečujú interlexikálnu dekompozíciu významu. CVAL a DVAL intralexikálnu. Ich striedavá aplikácia nám umožňuje kondenzovať vety do menných fráz, zložených a odvodených slov a spätne rozvíjať zloženiny na vety v plnom znení.

7.18 Analógia s chémiou

Tesniérovu analógiu (1957) s valenciou chemických prvkov môžeme plodne rozvinúť aj ďalej: Slová sú chemické prvky. Syntagmy a kompozitá predstavujú chemické zlúčeniny. Formálno-sémantické príznaky sú elektróny a voľné miesta pre ne vo valenčných sférach prvkov (slov). Medzi nimi pôsobia príťažlivé a odpudivé sily, ktoré umožňujú či zabraňujú aby sa prvky zlučovali s rôznou väzobnou silou. Petrifikované frazeologické štruktúry sú inertné zlúčeniny. Zlučovanie závisí v prvom rade od zúčastnených prvkov (lexikálnych jednotiek), ale niekedy aj od tretieho prvku, ktorý pôsobí ako katalyzátor.

Inventár elektrónov (príznakov) prvku (slova) definujú slovníky. Príťažlivé a odpudivé sily medzi nimi definujú kolokačné funkcie.

8.

KONTEXTOVÁ ANALÝZA

Kontextová analýza zisťuje, akým spôsobom sa frázy vo vete vzťahujú na objekty reálneho sveta. Znamená to priradiť frázy jazyka k reálnym objektom vonkajšieho sveta, resp. popriraďovať k sebe tie frázy, ktoré sa vzťahujú na jeden a ten istý vonkajší objekt.

Máme tri základné druhy referenčných vzťahov:

Referencia	anaforická	(odkazovanie na entity v predch. texte)
	kataforická	(odkazovanie na entity v nasled. texte)
	exoforická	(ukazovanie na entity vonkajšieho sveta, nespomenuté v texte)

Analýza exoforického ukazovania vyžaduje, aby počítač obsahoval popri jazykovom modeli ešte jeden paralelný model vonkajšieho sveta, vyjadrený v nejakom inom symbolovom systéme (napr. obrazový materiál).

SAPFO takýto model vonkajšieho sveta neobsahuje. Obmedzíme sa iba na jednoduché anaforické odkazovanie, ktoré môže byť dvojaké:

anaforická referencia	vnútrovetná
	medzivetná

Na rezolvenciu anafor budeme aplikovať kombináciu niekoľkých jednoduchých metód prevzaných z literatúry. Vykonať kontextovú analýzu je vo všeobecnosti veľmi ťažká úloha. Vyžaduje kombinovať viac metód, brať ohľad na viacero rôznorodých faktorov - a aj potom je výsledok zaručený iba čiastočne.

8.1 Príklady koreferencie

Odkazovanie je možné tak na menné, ako aj na slovesné frázy. Koreferujúce frázy sú označené rovnakým indexom:

Jano$_i$ spoznal seba samého$_i$. *(anafora)*
*Stretol som Jana$_i$ a pozdravil som **ho**$_i$.*
*Našiel som modré pero$_i$ a červenú ceruzku. **Pero**$_i$ bolo zlomené.*
*Závod vyhral$_i$ voz zo stajne Ferrari. Luigiho **to**$_i$ prekvapilo.*

*Je to **on**$_i$ - cigán$_i$, černý cigán$_i$ ...* *(katafora)*

Referencia sa môže týkať aj objektov, ktoré v predchádzajúcom texte neboli spomenuté samostatne. Môžu byť prvkom viacprvkovej množiny, časťou celku alebo v nejakom inom vzťahu k skôr spomenutej entite:

Vybral som jednu košeľu$_i$. *Mala odtrhnuté* ***gombíky***$_{par\ i}$
Kúpili sme dve gitary$_i$. ***Menšia gitara***$_i$ *bola zo smrekového,* ***väčšia***$_i$ *z mahagónového dreva.*
Mám dve deti$_i$. ***Syn***$_i$ *má osem rokov,* ***dcéra***$_i$ *dvanásť.*

8.2 Zoznam konštituentov

Nominálne frázy buď identifikujú nejaký starý objekt v kontexte už spomenutý, alebo uvádzajú objekt úplne nový.

Kontextový analyzátor používa tzv. *zoznam konštituentov* (history list). Je to usporiadaný (prípadne aj ohodnotený) zoznam objektov, ktoré sa spomínali v posledných niekoľkých vetách.

Ak analyzovaná fráza ukazuje na úplne nový objekt, pridá sa tento objekt do zoznamu konštituentov. Ak odkazuje na objekt už spomenutý, hľadá sa v zozname konštituentov, ktorý to je. Novoprichádzajúca informácia sa potom pripája k informácii, ktorá sa v databanke o tomto objekte už nachádza.

8.3 Syntaktická forma (určitosť, zvratnosť a vlastné mená)

Jedným z oporných bodov pre riešenie kontextovej koreferencie sú syntaktické kategórie apelatívnosti, zvratnosti a určitosti.

Propriálnosť. Kontextová analýza vlastných mien je relatívne jednoduchá. Ukazujú priamo na jednoznačne určenú osobu či objekt. Ak sa v písanom texte vyskytnú viaceré objekty s rovnakým menom, bývajú rozlíšené nejakým prívlastkom, aby sa neplietli.

Janko, Marienka, Karol IV., Alexander Veľký, Jozef Arimatejský, Južná Amerika

Reflexivita. Zvratné zámeno sa vzťahuje na subjekt predchádzajúcej vety, nezvratné zámená na ostatné konštituenty.

Jano$_k$ *stál za Ferom*$_i$. *Fero* ***sa***$_i$ *videl v zrkadle.*
Fero ***ho***$_k$ *videl v zrkadle.*

Určitosť. Neurčitá menná fráza uvádza na zoznam konštituentov nový objekt. Určitá menná fráza odkazuje na objekt skôr spomenutý.

V analytických jazykoch je kategória určitosti/neurčitosti gramatikalizovaná v podobe *určitého* a *neurčitého* člena. Aj vo flektívnych jazykoch sa dá použiť, ale iba ako doplnkový faktor. Kategóriu určitosti/neurčitosti možno „vyčítať" zo zámen v prípade, že menná fráza je nimi rozvitá:

nejaký pes	(a dog, ein Hund, un cane)
ten pes	(the dog, der Hund, il cane)

Využitie uvedených troch faktorov možno zhrnúť do tabuľky, platnej pre anglický ale aj pre slovenský jazyk (Allen, 1987):

Forma	Medzivetná anafora	Vnútrovetná anafora	Neanaforické ukazovanie
zvratné zámeno	áno, subjekt	nie	nie
nezvratné zámeno	áno, nie subjekt	áno	zriedka
určitá nf	nie	áno	áno
neurčitá nf	nie	nie	áno
vlastné meno	nie	nie	áno

8.4 Kategoriálne zhody (osoba, rod, číslo, pád, životnosť)

Vo flektívnych jazykoch máme pri kontextovej analýze možnosť výdatne využívať aj *zhodu v gramatických kategóriách*. Ak má nejaké zámeno alebo menná fráza odkazovať na inú frázu, musia sa zhodovať v rode, čísle, životnosti, prípadne aj v osobe a v páde:

Janko$_i$ chcel ísť s Marienkou$_k$ do lesa na maliny, ale mal$_i$ prácu.
...ale mala$_k$ prácu.

8.5 Stupeň aktivácie

Uvedené syntaktické podmienky pomáhajú obmedziť množinu referentov prichádzajúcich do úvahy, ale nie vždy zaručujú jednoznačnú analýzu. Môže sa stať, že napríklad k nejakému neurčitému zámenu nám ešte stále zostane viac menných fráz, ktoré sa s ním zhodujú vo všetkých kategóriách a prichádzajú do úvahy ako potenciálni kandidáti.

V takom prípade rozhoduje poradie na zozname konštituentov. Intuitívne sa neurčité zámeno bude vzťahovať s oveľa väčšou pravdepodobnosťou na nedávno spomenuté frázy, než na frázy spomenuté v kontexte dávnejšie. Toto pravidlo však tiež neplatí vždy:

Peter$_i$ navštívil Pavla$_k$. Potom sa (Peter$_i$) vrátil domov.

E. Hajičová a kol. (1990) navrhli algoritmus, ktorý každému prvku v zozname konštituentov priraďuje jeho *stupeň aktivácie*. Čím väčší stupeň aktivácie prvok má, tým pravdepodobnejším kandidátom na referent sa stáva. Pre stupeň aktivácie je rozhodujúce poradie, ale aj dôležitosť syntaktickej pozície, v ktorej bol objekt spomenutý, prítomnosť v tematickej alebo ohniskovej časti vety, opakovanie referencie pomocou dôrazového alebo nedôrazového zámena a i.:

(i) prvky stojace v ohnisku vety, referované podstatným menom alebo dôrazovým zámenom, dostanú maximálny stupeň aktivácie (MAX);

(ii) prvky referované podstatným menom v *téme* vety sa aktivujú o jeden stupeň nižšie (MAX-1) od prvkov v ohnisku;

(iii) pronominálna referencia na nejaký prvok v tematickej časti vety udržuje stupeň aktivácie nezmenený;

(iv) aktivácia prvkov, ktoré sa vo výpovedi nespomínajú, poklesne o dva stupne.

8.6 Inferenčná metóda

Posledná možnosť je rozhodnúť medzi rovnocennými kandidátmi inferenciou nad bázou poznatkov. Táto možnosť sa dá využiť, ak ide o frázu rozvitú prívlastkami.

Majme napr. frázu *hnedý kôň na lúke*. Nech sa v databanke a na zozname konštituentov nachádzajú štyri kone:

kôň$_1$	kôň$_2$	kôň$_3$	kôň$_4$
«AGS jazd$_1$-	LOK» stajň$_4$-	ATR» hned$_1$-	ATR» čiern$_1$-
VIA» lúk$_1$-		LOK» lúk$_2$-	
		«PRO pás$_1$- sa	

Treba rozhodnúť, o ktorého z už predtým spomínaných koní ide. Budeme hľadať referent s indexom *i*, ktorý vyhovuje rovnici:

kôň$_i$-
ATR» hned$_j$-
LOK» lúka$_k$-

Referencia subkonštituentov sa rieši rekurzívne. Ak po unifikácii rámca s bázou poznatkov vyjde ako riešenie jediná n-tica objektov, ide o hľadané referenty. V našom prípade sú to *kôň*$_3$, hnedý$_1$ a *lúka*$_2$. Ak je fráza s bázou poznatkov unifikovateľná viacerými spôsobmi, veta zostane kontextovo viacznačná.

8.7 Algoritmus SAPFO

1. Ak ide o neurčitú nf - vytvor nový referent.
2. Ak ide o zvratné zámeno - referent zreťaz s vetným subjektom.
3. Ak ide o nezvratné zámeno - testuj všetky predchádzajúce menné frázy na zhodu v gramatických kategóriách. Najprv vnútrovetne, potom medzivetne.
4. Ak ide o určitú mennú frázu - hľadaj medzivetnú anaforu s najväčším stupňom aktivácie.
5. Inferenciou vyber z množiny kandidátov - ak ich zostalo viac ako jeden.

8.8 Problémy

Sebareferencia

*The man, who dropped **his** hat.*

Zámeno *his* (jeho) odkazuje na objekt, v ktorom je samo obsiahnuté. To znamená, že by muselo odkazovať na frázu, ktorá sa v zozname konštituentov nemôže nachádzať, lebo jej analýza ešte nebola ukončená.

Odkazy na súvzťažné objekty.

PAR: *Píše tá ceruzka$_i$? Nie, má zlomenú **tuhu**$_{PAR\ i}$.*
LOK: *Vstúpil som do kuchyne$_{LOK\ i}$. **Sporák**$_i$ nehorel.*
DON: *Keď sme kupovali$_i$ mlynček, **predavač**$_{DON\ i}$ nám ho predviedol.*

S príchodom konštituentu by sme do zoznamu konštituentov museli zaradiť aj zoznam všetkých s ním rolovo súvisiacich objektov. Každú novú mennú frázu by sme potom preskúšali aj na nesynonymné vzťahy k predošlým objektom v zozname.

Avšak vzťahy k objektom v kontexte môžu byť aj viacstupňové:

*Keď sme vstúpili do kuchyne$_{lok\ koh\ i}$, videli sme, že unikal **plyn**$_i$.*

plyn
KOH: sporák
LOK: kuchyňa

Množinová referencia.

*Pri vykopávkach sa našli 3 amfory, hrdzavý nôž a 20 zlatých mincí. **Tieto artefakty** sú uložené v budove národného múzea.*

*Pri prvom ponorení sme našli sedem strieborných toliarov. **Najstarší** bol datovaný 1793.*

Objekty v zozname fráz by sa museli dať kombinovať do množín. Budeme utvárať všetky možné podmnožiny ? (2^n)

Nešpecifickosť jazyka. Rámcovosť reprezentácie.

Päť chlapcov zjedlo desať pomarančov.

Ktorý z chlapcov zjedol koľko pomarančov ? Prirodzený jazyk má tú vlastnosť, že jednoznačne špecifikuje iba tie vzťahy, ktoré sú podstatné z hľadiska komunikačného zámeru. Ostatné môžu zostať bližšie nešpecifikované, hmlisté. Vždy ich možno špecifikovať dodatočne - ak je to potrebné.

Táto vlastnosť je skôr výhodou, ako nevýhodou prirodzeného jazyka. Preto si myslíme, že aj formálna reprezentácia by mala dokázať reprezentovať vzťahy špecifikované len rámcovo. Je to jedna zo slabín všetkých reprezentácií orientovaných na matematickú logiku: ťažko pracujú s neurčitosťou (Kamp, 1989).

Atributívne a referenčné frázy

Frázy môžu byť dvojakého typu: *referenčné a atributívne*. Tieto dva druhy fráz nemožno miešať, inak vzniknú nekorektné sylogizmy:

1. Jozef je predseda klubu.
2. Chcem byť predseda klubu.
3. Chcem byť Jozef.

Negácia

Kedy je referencia korektná v spojení s negáciou ?

*Nemám psa. Ukradli mi **ho**.*	... PES17
**Nemám psa. Je to foxteriér.*	... ?

Kvantifikácia

Každý chlapec dostal darček.

Tu nejde len o jednoduché priradenie, ale o celú n-ticu referenčných priradení. Chlapci$_{1,2...n}$ dostali darček$_{1,2...n}$, každý ten svoj.

Generickosť

Smreku sa darí vo vysokých polohách.
Levy sú nebezpečné.

V prvej vete nejde o jeden konkrétny smrek, ale o smrek ako druh, t. j. o všetky smreky. V druhej vete tiež nejde o konkrétnu množinu levov, ale o levy ako také.

9.

FRAZEOLÓGIA

„*Frazeologická jednotka* (frazeologizmus) je ustálené slovné spojenie, ktoré sa vyznačuje obraznosťou a nerozložiteľnosťou svojho významu, ako aj expresívnosťou" (Mlacek, 1977).

Frazeologizmy predstavujú veľmi disparátny inventár javov a vyžadujú si diferencovaný prístup. Čo do významu aj formy sa správajú špecificky. V tejto kapitole popíšeme pokus o reprezentáciu a začlenenie frazeologických jednotiek do formálneho modelu jazyka a problémy, ktoré s tým súvisia. Niektoré triedy frazeologizmov sa darí dobre formálne reprezentovať, niektoré iba čiastočne a na formalizáciu niektorých sme museli rezignovať.

Tvorba strojového frazeologického slovníka vyžaduje manuálny a jednotlivý prístup. Slovník umožňuje frazeologickému modulu nahrádzať frazémy ich doslovnými významami a naopak. Tým vznikajú parafrázy:

Ondrej zavesil remeslo na klinec.
Ondrej sa prestal zaoberať remeslom. Fra^{-1}

Viacslovné ekvivalenty. V skorších teóriách (Vinogradov, 1947, Bally, 1961) sa v zásade počítalo s ekvivalenciou frazeologickej jednotky a slova. Nie všetky frazeologizmy sa však dajú parafrázovať jedným slovom (*Padla kosa na kameň*). SAPFO dovoľuje reprezentovať frazeologizmy synonymné s jednoduchou lexémou (*od buka do buka - nepresne, vyjsť na žobrácku palicu - schudobnieť*), ale pripúšťa aj ekvivalenciu s komplexnými syntagmami a vetami:

frazeol (ani_do_súdneho_dňa,nikdy).
frazeol (vo_dne_v_noci,neustále).

$$\text{frazeol}\ \left(\begin{bmatrix}\text{zavesiť} \\ \text{PAC» X} \\ \text{DES» klinec}\end{bmatrix},\ \begin{bmatrix}\text{zaoberať sa}^{+fin} \\ \text{PAC» X}\end{bmatrix}\right).$$

$$\text{frazeol}\ \left(\begin{bmatrix}\text{spadnúť}^{+neg} \\ \text{PAC» koruna} \\ \text{ORI}_g\text{» hlava}\end{bmatrix},\ \begin{bmatrix}\text{stratiť}^{+neg} \\ \text{PAC» dôstojnosť} \\ \text{+hovor}\end{bmatrix}\right).$$

Spôsob reprezentácie závisí od stupňa ustálenosti frazeologickej jednotky. Niektoré frazeologizmy reprezentujeme sémantickou sieťou, iné ako reťazec znakov. Alebo kombinovane: zmrazené časti frazeologizmu hotovým reťazcom foném,

variabilné časti sieťou, ktorá umožňuje pohyblivosť, ohybnosť (porov. aj Filipec - Čermák, 1985).

9.1 Petrifikované a voľné frazeologické jednotky

Petrifikované frazeologické jednotky možno reprezentovať ako jednoliaty reťazec znakov, ktorý sa neanalyzuje, ale vystupuje ako samostatná lexikálna jednotka. Jednotka preberá syntaktické funkcie lexikálnej jednotky, s ktorou je synonymná:

frazeol(do_chlapa,všetci).
frazeol(ani_zbla,ani trochu).
frazeol(ani_za_mak,ani trochu).

Môže byť synonymná aj s viacerými syntaktickými druhmi zároveň:

frazeol(na_jedno_kopyto,rovnaký).
frazeol(na_jedno_kopyto,rovnako).

V mnohých prípadoch ide o frazeologické jednotky, ktoré by ani sémantickou sieťou v našom formalizme reprezentované byť nemohli. Môžu obsahovať zastarané, defektné až neanalyzovateľné morfologické tvary a syntaktické formy, alebo dovoľujú jedinečnú, inak neprípustnú sémantickú spájateľnosť. Neprešli by teda cez syntakticko-sémantickú analýzu. (*My o vlku a vlk za humny; ani zbla; jeden čihi, druhý hota; dve na tri; piate cez deviate; prísť do chlapa*).

Voľné frazeologické jednotky. Majú ustálenú lexiku, ale z hľadiska syntaxe a tvaroslovia sa správajú ako živé. Tieto naopak nemožno uchovávať vo forme reťazca znakov, pretože sa vyskytujú v rozličných formách. Vo frazeologickom slovníku sú uložené vo forme siete. Prechádzajú regulárnou syntaktickou a sémantickou analýzou a identifikujú sa až na úrovni hĺbkových štruktúr (pokusom o unifikáciu medzi kusmi sémantických sietí):

$$\text{frazeol}\left(\begin{bmatrix} \text{nosiť} & \\ & \text{OBJ» drevo} \\ & \text{DES}_1\text{» hora} \end{bmatrix}, \begin{bmatrix} \text{robiť} & \\ & \text{OBJ» X} \\ & \text{MDS» zbytočne} \end{bmatrix}\right).$$

Najväčší problém predstavujú *frazeologické jednotky, ktoré nie sú celkom petrifikované, ale ani celkom voľné*. Pripúšťajú tvarové varianty, ktorých paradigma je však neúplná. Frázy *labutia pieseň, mať zajačie úmysly, kameň úrazu, začarovaný kruh* vyžadujú ponechať skloňovanie, ale ak ich reprezentujeme sieťou, musíme nejako zabrániť, aby nevznikali formy *mať úmysly zajaca; úrazový kameň; kruh, ktorý niekto začaroval*.

Fráza *nepreložiť krížom slamy* dovoľuje časovať sloveso, ale vyžaduje dodržať negáciu aj partitívny genitív. *(Ne)hádzať flintu do žita* dovoľuje kladnú aj zápornú

formu slovesa. *Prísť o zuby* má frazeologický význam iba v pluráli. Spájateľnosť fráz *v Adamovom rúchu, v Evinom rúchu* je obmedzená čo do rodu; v prvom prípade na mužský, v druhom na ženský.

Vo frázach *tárať dve na tri, vyjsť na psí trisiatok, počuť na vlastné uši, vidieť na vlastné oči* hlavné slovesá pripúšťajú plnú časovaciu paradigmu (*videl, uvidíme, vidím na vlastné oči)* ale časti *dve na tri, na psí tridsiatok, na vlastné oči,* sú ustálené. Môžeme ich reprezentovať kombinovane takto:

$$\text{frazeol}\left(\begin{bmatrix}\text{báchať} \\ \text{RDET»dve\_na\_tri}\end{bmatrix}, \begin{bmatrix}\text{tárať} \\ \text{+expr}\end{bmatrix}\right).$$

$$\text{frazeol}\left(\begin{bmatrix}\text{vyjsť} \\ \text{RDET»na\_psí\_tridsiatok}\end{bmatrix}, \begin{bmatrix}\text{schudobnieť} \\ \text{+expr}\end{bmatrix}\right).$$

Reprezentácia čiastkových paradigiem gramatických kategórií by si vyžadovala zaviesť len kvôli frazeologizmom osobitnú notáciu a aj osobitný mechanizmus, ktorý by ju potom spracúval. To však pokladáme za neprirodzený zásah do jednoliatej elegancie celého systému.

Čiastkové gramatické paradigmy teda reprezentujeme tak, že zavádzame niekoľko sémantických pádov, ktorých sémantikou je, že nie sú sémantické, ale syntaktické. Oboje sú však spracúvané jedným a tým istým unifikačným formalizmom. Napríklad LDET je taký pád, ktorý sa vždy prejaví len vo forme ľavostranného zhodného prívlastku. NA+LOK je pád, ktorého povrchové formy sú obmedzené len na predložkový pád na+lokál. LDET, NA+LOK neunifikujú so žiadnym iným sémantickým pádom, nevstupujú teda do transpozícií, ale miesto svojho výskytu petrifikujú len na jedinú špecifickú syntagmu. Význam LDET, NA+LOK nie je špecifikovaný, čo je tiež správne, lebo ide iba o formu:

$$\text{frazeol}\left(\begin{bmatrix}\text{mať} \\ \text{OBJ: úmysel}^{+pl} \\ \text{LDET: zajac}\end{bmatrix}, \begin{bmatrix}\text{utiecť}^{+volun} \\ \text{+colq}\end{bmatrix}\right).$$

$$\text{frazeol}\left(\begin{bmatrix}\text{byť} \\ \text{BEN» X} \\ \text{OBJ» Y} \\ \text{PROTI+DAT» srsť-}\end{bmatrix}, \begin{bmatrix}\text{páč- sa}^{+neg} \\ \text{BEN» Y} \\ \text{OBJ» X}\end{bmatrix}\right).$$

9.2 Reprezentácia ďalších frazeologických javov

Fakultatívne členy. Niektoré frazeologizmy sa vyskytujú vo viacerých variantoch; pripúšťajú napr. fakultatívne členy, ktoré možno zamieňať či elidovať (*ako blesk [hrom] z jasného [čistého] neba*). Každý variant možno reprezentovať osobitne.

Alternanty. Frázy, ktoré obsahujú syntakticky povinné, ale lexikálne voľné členy (tzv. alternanty), možno dobre reprezentovať pomocou zreťazenia voľných premenných v PROLOGu:

$$\text{frazeol}\left(\begin{bmatrix}\text{vziať si}\\ \text{OBJ» X}\\ \text{MDS» do\_hlavy}\end{bmatrix},\begin{bmatrix}\text{zaumieniť si}\\ \text{OBJ» X}\end{bmatrix}\right).$$

Ustálené prirovnania. V súlade s J. Branickou (1991) sme frazeologické komparatíva riešili ako súčasť paradigmy *magnoným* v rámci lexikálno-paradigmatického systému (*krokodílie slzy, mlčať ako hrob*). Ale viacslovné ustálené prirovnania so zložitejšou štruktúrou možno tiež zaradiť do frazeologického slovníka (*chudobný ako kostolná myš, hladný, že by klince pojedol, boj(ovať) na život a na smrť, podobať sa ako vajce vajcu*).

Paremiologické útvary. Príslovia, porekadlá a úslovia majú spravidla ucelenú a dokonale ustálenú vetnú formu. Parafrázovač si ich môže zapamätať, no nie parafrázovať ich význam (*Darovanému koňovi nehľaď na zuby. Čo na srdci to na jazyku. Jedna lastovička nerobí leto*). Informácia v slovníku teda poslúži iba na to, aby sa program o doslovnú interpretáciu ani nepokúšal.

Paradigmatické vzťahy. Frazeologické jednotky môžu vstupovať do paradigmatických vzťahov paralelných so vzťahmi v slovnej zásobe. Ide najmä o synonymiu a antonymiu, ktorá môže byť v slovníku taktiež zachytená. (synon. *nepreložiť krížom slamy - nepohnúť ani prstom - chodiť so založenými rukami - ležať hore pupkom*, anton. *mlátiť prázdnu slamu - trafiť klinec po hlavičke*).

Polysémia. Pri frazeologických jednotkách je zriedkavá, no môže sa vyskytnúť. Napr. E. Smiešková (1989) uvádza pre frázu *trafila (padla) kosa na kameň* dva významy: *a) stretli sa rovnako rozhodní, neústupčiví ľudia, b) nastali prekážky, objavil sa odpor*.

Štylistická hodnota frazeologizmov. Frazeologické jednotky môžu byť príznačné z hľadiska nejakého štýlu, nárečia alebo svojou expresívnosťou, zosileným významom. Štylistický príznak sa po transformácii stáva súčasťou sémantickej siete.

Iné. Niektoré slovné spojenia pripúšťajú aj doslovnú aj prenesenú interpretáciu. Rozlíšiť medzi nimi je ťažké. Rozhodujúci môže byť kontext. Fráza *indiánske leto* by sa v reportáži o Kanade chápala doslovne, na Slovensku by bola ekvivalentná s frazeologickou jednotkou *babie leto*.

Navyše doslovný ekvivalent, ktorý uvádzame ku každému frazeologizmu ako jeho synonymum, býva nezriedka iba jedným z možných a silno zužuje pôvodný význam, ktorý tým stráca na farbitosti a výrečnosti.

10.

ŠTYLISTIKA

Štylizácia spočíva vo výbere prostriedkov pre slovesný prejav. Pretože úlohou SAPFO je tvoriť rodiny čo najrozmanitejších parafráz zviazaných jedným obsahom, je to štylistický systém par excellence.

Parafrázovač disponuje arzenálom vyjadrovacích prostriedkov na všetkých úrovniach jazyka (pozri kap. 13 o algebre syntakticko-sémantických manipulácií s jazykom). Množstvo parafrázovacích prostriedkov dáva priestor pre voľbu v medziach gramatickosti a formálno-sémantickej kompatibility.

10.1 Štylistický modul

Povedomie o štylistickej konzistencii textu je stelesnené v *štylistickom module* programu. Štylistický modul udržuje kontrolu nad všetkými procesmi z hľadiska štýlotvorného. Tým zamedzuje vzniku štylisticky disparátnych viet *(*Otče, tys' ohákol sa ako chuliš! *Rty rudé sťa sadajúce slnce a oči circa ako MUDr. Pšenička z ÚPF NK...)*

SAPFO umožňuje:

a) počas analýzy identifikovať príslušný štýl a nárečie prejavu, a kontrolovať ich dodržiavanie;

b) počas syntézy zvoliť jeden zo štýlov (hovor., umel., odb., admin., kniž., arch.) alebo jedno zo slovenských nárečí (liptov., šariš.), resp. spisovný jazyk v neutrálnom štýle a konštruovať parafrázy v súlade s touto voľbou.

Štylistické prostriedky systému sú lexikálne, hláskoslovné, tvaroslovné, slovotvorné a syntaktické. Z nich vždy iba tie, ktoré sa nám podarilo formalizovať.

Pretože SAPFO nepracuje s textami, ale iba s jednotlivými vetami a súvetiami, najviac dvojicami viet, štylistické možnosti systému sú obmedzené na mikrokompozíciu vety. Makrokompozícia by vyžadovala komplexné výpočty nad veľkými textovými celkami. Formalizáciu takého procesu si zatiaľ nevieme predstaviť.

K úplnosti celého modelu chýba tiež zvuková zložka jazyka. Rozšírenie o zvukovú stránku by si vyžiadalo zavedenie intonačnej notácie a zadávanie vstupných viet vrátane intonačných značiek.

10.2 Štylistické prostriedky lexikálne

Dominantným štylististickým prostriedkom je výber lexiky. Potrebnú informáciu poskytuje štylistický slovník; obsahuje štylisticky príznakové slová spolu s ich

štylistickými príznakmi. Parafrázy vznikajú jednoduchou zámenou lexém, pričom k sémantickým komponentom druhej vety pribudne príslušný komponent štylistickej príznačnosti:

Luna lká = Mesiac plače +2poet

Štylistický modul pritom kontroluje, či využitie tých-ktorých lexikálno-sémantických parametrov je pri danom nastavení štýlu prípustné. Slová s príznakom expresívnosti nedovolí použiť v exaktných štýloch. Hovorové slová sa nedajú použiť v knižnom štýle a naopak. Zo spisovného jazyka vylúči dialektizmy. Symbolizáciu, magnonymizáciu, metaforické frazeologizmy a prirovnania dovolí používať v umeleckom, ale nie v odbornom a administratívnom štýle, v ktorých naopak dovolí skratky a značkové symboly, ale nie poetizmy alebo citoslovcia.

Pri deminutívach program nedokáže rozlíšiť, či ide o deminutívnu minorizáciu, alebo eufemizáciu. Slovo *koník* môže označovať koňa, ktorý je fyzicky malý, alebo koňa, ku ktorému máme citový vzťah. Výsledok štylistickej analýzy zostáva dvojznačný:

mesiačik = $mesiac_1$ +cit	(nebeské teleso)
mesiačik = $mesiac_2$ +min	(mesiačik pomaranča)

V iných prípadoch môže naopak práve štylistický modul prispieť k zjednoznačneniu frázy. V súčinnosti (synergii) so syntaktickým a sémantickým modulom môže mať rozhodujúce slovo pri riešení polysémie. Napríklad slová *vyletieť* (z hniezda, zo skúšky) a *zašiť sa* (do vreca, skryť sa) majú po dva významy, z ktorých jeden je spisovný a druhý slangový. Ak štylistický modul počas analýzy už identifikoval celkový štýl frázy, význam polysemického slova sa selektuje v súlade s celkový štýlom.

10.3 Štylistické prostriedky hláskoslovné

Hláskolovie je štylisticky príznakové najmä v súvislosti s dialektmi. Do fonologického modulu pribudnú ďalšie definície hlások a ich elementárnych vlastností, ktoré sú platné pod podmienkou, že dialekt je nastavený na určitú hodnotu:

```
konson([ć, dź, ś, ź]) :- dial(spiš).
vokál([u̯]) :- dial(spiš).
vokál([ä]) :- dial(spiš),!,fail.
párové(s, ś) :- dial(spiš).
diftong([u̯o, i̯e]) :- dial(orava).
vokál([a]) :- dial(orava).
```

Ak je nastavený napr. šarišký dialekt, hláska *s* sa počíta za konsonant. Inak sa počíta za neexistujúcu. Prestavením dialektu sa teda mení fonologický inventár systému.

10.4 Štylistická hodnota gramatických tvarov

Nielen slová, ale aj časti slov môžu byť nositeľom štylistickej hodnoty: niektoré tvarotvorné a slovotvorné afixy alebo syntaktické konštrukcie. Štylisticky zafarbená môže byť *forma* ako taká. Tu platí pravidlo, že štylisticky príznakové zafarbenie rastie s klesajúcou frekvenciou jazykovej formy.

Slovotvorba. Napríklad slovotvorné prípony *-oš, -úň, -áň, -aj, áč, -an* obsahujú pejoratívny odtienok (*pijan, chvastúň, frfloš, brucháč, táraj*). Prípony *-ečko, -očko, -ček, -ička* nesú príznak deminutívnosti alebo laudatívnosti, *-izeň, -írovať* (*dedovizeň, mašírovať*) znejú archaicky, hromadné prípony *-ie, -ina, -ež, -ač, -stvo* (*Slovač, človečenstvo*) sa uprednostňujú v rečníckom štýle, adverbiálna prípona *-ačky* (*ležiačky, kolenačky*) funguje v hovorovom štýle.

Dištinktívnosť prefixov zo štylistického hľadiska sa ukazuje až pri frekvenčnej analýze (Mistrík, 1977); v rámci vety teda nemá význam.

Tvaroslovie. Z tvaroslovia sú štylisticky príznakové niektoré nízkofrekventované tvarotvorné postupy. Napríklad vokatív je príznačný archaickosťou (*bože, otče*), sčasti aj mužský životný datív a lokál utvorený príponou *-u* (*bohu, otcu*). Expresívny je genitív *nemám matky* (oproti *nemám matku*). Štylisticky diferencované môžu byť dvojtvary (*dievčatá; dievčence, hov.*).

Expresívnu funkciu majú niektoré príklonky a predpony zámien (*dajže, ktože, komuže, kdeaký, kadejaký*). Predpony *da-, voľa-* sú hovorové, *kde-, kade-, poda-, leda-, bohvie-, čertvie-* sú expresívne.

Prechodníky a príčastia sú typické pre knižný štýl.

10.5 Štylistické návestia. Dialekty a archaizmy

Štylistický modul kontroluje skladbu slov a viet tak, že na všetkých miestach programu, kde sa utvárajú štylisticky príznakové formy, má svoje návestia. Štylistické návestia sú súčasťou podmienok a akcií prechodových sietí syntaktického modulu, skloňovacích a časovacích vzorcov morfologického modulu, kolokačných funkcií (LVAL, RVAL) sémantického modulu ako aj slovotvorných vzorcov (CVAL, DVAL) slovotvorného modulu. Na každom mieste, kde sa teda tvorí štylisticky príznaková forma, sa vždy aktivuje aj štylistické návestie - a príslušnú gramatickú formu, syntaktickú konštrukciu, slovotvornú deriváciu alebo sémantickú kolokáciu dovolí utvoriť iba so súhlasom štylistického modulu. Napríklad tvorbu substantívnych derivátov sufixom *-ce* (*slnce, drievce*) možno podmieniť nastavením štýlu na archaický:

DVAL(km(sub,konkr,ø,-ce),ko(sub,konkr)) :- štýl(arch).

Alebo tvorbu prechodníkov podmieniť knižným štýlom:

tranz(Koreň,Tvar) :- gen_tranz(Koreň,Tvar),štýl(kniž).

Dialekty. Morfologický modul možno opatriť návestiami, ktoré usmernia tvorenie tvarov v súlade s niektorým dialektom. Napríklad

```
subst(Pa,pl,Rod,_,och) :- (Pa=gen;Pa=ak;Pa=lok),dial(spiš).
subst(dat,pl,Rod,_,om) :- dial(spiš).
```

zavedie nové skloňovanie v genitíve, akuzatíve, lokále a v datíve plurálu substantív bez ohľadu na rod (*chlopom, o chlopoch, ženom, o ženoch, mestom, o mestoch)*, ktoré bude fungovať iba s podmienkou, že dialekt je nastavený na spišský, inak nie je aktívne. Podobne môžeme zadefinovať napríklad východoslovenský l-ový slovesný tvar a infinitív (*robiľi, robic)* alebo oravské skloňovanie a časovanie (*chlapou, rodiča, srca, dobruo, robiu)*:

```
lpar(Rod,pl,ľi) :- dial(spiš).
infin(c) :- dial(šariš).
subst(nom,pl,str,srce,ã) :- dial(orava).
subst(ak,pl,muž,chlap,ou̯) :- dial(orava).
adj(nom,sg,str,poz,dobr,u̯o) :- dial(orava).
lpar(muž,sg,u̯) :- dial(orava).
```

Môžeme oživiť aj archaické skloňovania (*s kosťu, o kosťech)*, duál (*očima, rukama*), aorist (řesevě):

```
subst(ins,sg,žen,kosť,u) :- štyl(16.st).
subst(lok,pl,žen,kosť,ech) :- štyl(16.st).
subst(ins,duál,žen,_,a) :- štyl(13.st).
aorist(1.os,duál,asigm,ě) :- štyl(12.st).
```

10.6 Štylistická hodnota syntaktických konštrukcií

Horeuvedeným spôsobom sú štylisticky ohodnotené aj syntaktické konštrukcie. ATN-sieť SAPFO má napríklad miesto aj pre gramaticky nadbytočné členy, avšak využitie takejto syntaktickej pozície spôsobuje pripojenie sémantického komponentu expresívnosti. Napríklad nadbytočné osobné zámená v nominatíve a datíve (citový datív):

Idem už preč → Idem ***ja*** *už preč, +expr.*
To je taký človek → To ***ti*** *je taký človek, +cit.*

Inverzia nezhodných prívlastkov je špecifikom poetického štýlu (*neba sklon, srdca hlas)*. Narušenie hierarchie zhodných prívlastkov je tiež príznakom emfázy.

Asyndetické spájanie viet v súvetí (čiarkou, bez spojky) nie je vhodné v exaktných štýloch, ktoré vyžadujú čo najväčšiu presnosť.

To sú niektoré štylistické možnosti syntaxe. Väčšina ostatných sa vymyká pokusu o formalizáciu.

11.

STAVEBNICOVÁ REPREZENTÁCIA VÝZNAMU

11.1 Vlastnosti sémantickej reprezentácie

[1] DEKLARATÍVNOSŤ - reprezentácia musí byť vyjadriteľná v nejakom symbolovom systéme a tak fyzicky realizovateľná v počítači.

[2] UNICITA - jedno-jednoznačné priradenie medzi významami a symbolmi. Dve vety musia mať rovnakú (podobnú) reprezentáciu práve vtedy, keď majú rovnaký (podobný) význam. Rozličné významy (čítania) tej istej vety majú rozličné reprezentácie.

[3] EXPLIKAČNÁ SILA - reprezentácia musí byť schopná vhodným spôsobom vyjadriť to, čo intuitívne nazývame „významom", a to pre čo najširšiu škálu viet jazyka.

[4] PROCEDURÁLNOSŤ - reprezentácia musí byť algoritmicky vyrobiteľná a tiež efektívne a zmysluplne manipulovateľná pri inferencii.

11.2 SRL (Sentence Representation Language)

SRL je symbolický jazyk pre reprezentáciu vetného významu.

Je to stavebnica, ktorá pozostáva z troch druhov tehál: plnovýznamových slovných koreňov, sémantických pádov a komponentov významu. Z tejto stavebnice možno utvárať kusy sémantickej siete takto: vo vrcholoch grafu sú umiestnené plnovýznamové slovné korene, na hranách sú sémantické pády a ku každému vrcholu možno priradiť ešte množinu komponentov významu.

Korene. Slovník SAPFO obsahuje momentálne 2222 slov v 300 slovotvorných hniezdach. Nerozložiteľných plnovýznamových koreňov, ktoré môžu tvoriť vrcholy sémantickej siete, je teda 300.

Sémantické pády. Množinu sémantických pádov sme zadefinovali v odseku 5.3. Je ich 88. Stoja na hranách sémantickej siete.

Komponenty významu. Princíp vydeľovania komponentov významu kontrastívnou diferenciáciou sme vysvetlili v odseku 0.16. SAPFO používa 66 komponentov významu.

11.3 Komponenty významu

smerovanie z	+ex	vysypať = sypať +ex +perf
od	+ab	odplaziť sa = plaziť sa +ab +perf
do	+in	vtekať = tiecť +in
na	+sur	našiť = šiť +sur +perf
pod	+sub	podliezť = liezť +sub +perf
nad	+supra	nadletieť = letieť +supra +perf
nahor	+alt	vzlietnuť = letieť +alt +perf
nadol	+deo	zletieť = letieť +deo +perf
k, pri	+ad	prisadnúť si = sadnúť si +ad +perf
za	+post	zašiel = šiel +post
pred	+ante	predísť = ísť +ante +perf
cez	+trans	precestovať = cestovať +trans+perf
okolo	+circ	obsekávať = sekať +circ
rozptýlene	+dsp	rozvešiavať = vešať +dsp
spolu	+con	zošívať = šiť +con
k výsledku	+rzt	vyšiť = šiť +rez +perf
fáza začiatok	+inc	rozsvietiť sa = svietiť +inc +perf
koniec	+fin	dofajčiť = fajčiť +fin +perf
súčasnosť	+sim	tancuje spievajúc = tancuje a spieva +sim
reflexívnosť	+ref	rozprávať sa = rozprávať +ref
zavŕšenosť	+perf	urobiť = robiť +perf
iteratívnosť	+iter	chodievať = chodiť +iter
distributívnosť	+dis	povhadzovať = vhodiť +dis
viacsmernosť	+dir	chodiť = ísť +dir
intencionálnosť	+itc	zlomilo sa mi pero = zlomil som pero -itc
intenzitnosť	+int	nabehať sa = behať +int +perf
kauzatívnosť	+caus	bieliť = belieť +kau
rezultatívnosť	+rez	zožltnutý = žltý +rez

dynamickosť	+dyn	belieť sa = byť biely +dyn
čas minulý budúci	+pret +fut	pršalo = prší +pret pokvitnú = kvitnú +fut
modálnosť vôľa schopnosť sklon zvyk možnosť nevyhnutnosť	+volun +abil +incl +usu +posib +neces	robiť +volun = chcieť robiť kujný = kuť +abil bojazlivý = báť sa +incl robí +usu = zvykne robiť robiť +posib = smieť robiť robiť +neces = musieť robiť
zápor	+neg	nezávislosť = závislosť +neg
kvantita	+pl	lístie = list +pl listy = list +pl
rod ženský mláďa	+fem +neu	pani = pán +fem psíča = pes +neu
miera malá zvýšená veľká maximálna	+min +maj +mag +max	červenkastý = červený +min poskočiť = skočiť +min kvietok = kvet +min krajší = pekný +maj prekrásny = krásny +mag tíško = ticho +mag stromisko = strom +mag najkrajší = pekný +max

kvantifikácia všeobecná existenčná	 +univ +exist	 všetky vtáky lietajú = vtáky lietajú +univ je aspoň jeden vták, čo nelieta = vták nelieta +exist

expresivita všeobecná sympatická antipatická	 +expr +cit +pej	 slniečko = slnko +cit žrať = jesť +pej
modálnosť opytovacia želacia rozkazovacia zvolacia podmieňovacia	 +rog +des +imper +excl +kond	 Príde Peter? = Peter príde. +rog Kiež príde. = Príde. +des Stáť! = stáť +imper Dúha! = dúha +excl robil by = robí +kond

štýl hovorový	+colq	auto = automobil +colq
novinársky	+jour	
archaický	+arch	žalár = väzenie +arch
umelecký	+poet	deva = dievča +poet
knižný	+lib	dievka = dievča +lib
odborný	+sci	vrabec poľný = poľný vrabec +sci
kancelársky	+admin	dolupodpísaný = podpísaný dole+admin
spisovnosť		
dialekt	+dial	ščiroch krídžial = štyroch krídel +gemer
úctivá forma		
vykanie	+rev	poďte = poď +rev
onikanie	+rev	ak dovolia = ak dovolíš +rev
dominantný člen	+dom	okná$^{+dom}$ domu = dom^{+dom} s oknami
ohniskový člen	+fok	príde Jano = Jano^{+fok} príde
zdôraznený člen	+ton	*Jano* príde = Jano^{+ton} príde

Spisovnosť, nocionálnosť, deklaratívna modálnosť, prítomný čas, nedokonavosť, singulár, mužský rod sú bezpríznakové charakteristiky.

12.

SYNTÉZA PARAFRÁZ

12.1 Intonácia

Najjednoduchšie parafrázy vznikajú zmenou intonácie. Presunom sémantického komponentu *+ton* posúvame intonačný dôraz na jednotlivé členy vo vete, ktoré potom syntaktický syntezátor zvýrazní kurzívou. Intonačne zdôraznené môže byť žiadne, jedno alebo viac slov, alebo niektoré časti slova:

ísť-$^{+ad\ +pret}$
PCS» sused-$^{+ton}$

Sused prišiel.

ísť-$^{+ad\ +pret\ +ton}$
PCS» sused-

Sused *prišiel*.

stúp-$^{+neg\ +ton\ +ex\ +pret}$
PCS» sused-

Sused *ne*vystúpil.

sused-
«PCS stúp-$^{+ex\ +ton\ +neg\ +pret}$
«PCS stúp-$^{+in\ +ton\ +pret}$

Sused ne*vy*stúpil, *na*stúpil.

12.2 Transfokácia

Transfokáciou presúvame ohnisko (rému) vnútri viet a fráz. Objektívne radenie vetných členov sme prevzali od E. Hajičovej a P. Sgalla (1987): 1. agens 2. adresát 3. paciens 4. pôvod 5. výsledok 6. spôsob 7. odkiaľ 8. nástroj 9. kam. Členy označené komponentom *+fok* sa oproti tomuto štandardnému poradiu presúvajú na koniec vety (frázy). Rematicky vyznačený môže byť žiaden, jeden alebo viac členov na každej z úrovní syntaktickej organizácie vety. Syntezátor podľa toho stvárňuje vetu slovosledne:

darov-$^{+neg\ +pret}$
DON» Jan-$^{+min}$
REC» Mári-$^{+min}$
PAC» kvet-$^{+pl\ +fok}$

Janko nedaroval Marienke kvety.

darov-$^{+neg\ +pret}$
DON» Jan-$^{+min\ +fok}$
REC» Mári-$^{+min}$
PAC» kvet-$^{+pl}$

Marienke nedaroval kvety Janko.

darov-$^{+neg\ +pret}$
DON» Jan-$^{+min}$
REC» Mári-$^{+min\ +fok}$
PAC» kvet-$^{+pl}$

Janko nedaroval kvety Marienke.

darov-$^{+neg\ +pret\ +fok}$
DON» Jan-$^{+min}$
REC» Mári-$^{+min}$
PAC» kvet-$^{+pl}$

Janko Marienke kvety nedaroval.

Transfokácia je možná aj vnútri nominálnych fráz. Avšak v 4.19 sme povedali, že poradie prívlastkov v nominálnej fráze je pevné; preradenie zhodných prívlastkov do postpozície nemá rematizačnú, ale štylistickú a polopredikatívnu funkciu (4.20, 10.6).

Napriek tomu sa transfokácia uskutočňuje, a síce voľbou valenčných funkcií počas formácie vety. Funkcia RVAL slúži ako rematizátor. Ak je niektorý člen označený ako ohniskový, syntezátor sa ho snaží umiestniť na pravej strane frázy a vyjadrí ho prostredníctvom funkcie RVAL. Tým sa miesta na pravej strane obsadia, čím ostatným členom zostane možnosť vyjadriť sa (prostredníctvom funkcie LVAL) na ľavej strane ako téma:

ísť-$^{+ad\ +fok}$
PCS» sused-

susedov príchod

vid-
KOG» Jan-$^{+min}$
FEN» dievč-$^{+fok}$
«AGS spiev-$^{+fok}$
ATR» pekn-

Janko vidí pekné dievča spievať.
Janko vidí pekné dievča ako spieva.

ísť-$^{+ad}$
PCS» sused-$^{+fok}$

príchod suseda

vid-
KOG» Jan-$^{+min}$
FEN» dievč-$^{+fok}$
«AGS spiev-
ATR» pekn-$^{+fok}$

Janko vidí spievajúce dievča, ktoré je pekné.

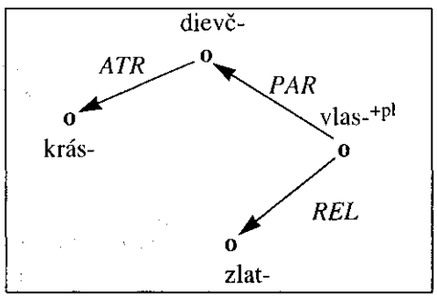
12.3 Transformácia

Trans*formáciou* rozumieme premenu *formy* pri zachovaní sémantických vzťahov. Typickými príkladmi transformácie sú transpozícia, nominalizácia a konverzia slovného druhu.

Formáciu kontrolujú valenčné funkcie. Sú to rovnice, ktoré definujú vzťah medzi významom a formou. Na každý sémantický pád v sémantickej sieti možno

aplikovať jednu z piatich valenčných funkcií: VAL, LVAL, RVAL, CVAL, DVAL. Ich rozličnými voľbami a kombináciami sa potom významové vzťahy stvárňujú do rozličných foriem.

VAL formuje aktanty do tvaru mennej frázy, predložkovej frázy, príslovky, infinitívnej konštrukcie a vedľajšej vety. Cez LVAL prechádzajú všetky zhodné prívlastky: adjektíva, číslovky, príčastia a zhodné doplnky. RVAL dáva participantom formu mennej, predložkovej alebo adjektívnej frázy, prístavku, príslovky, infinitívnej alebo prechodníkovej konštrukcie, vedľajšej vety a parentézy. Ak zveríme nejaký sémantický vzťah funkciám CVAL, DVAL, pokúsia sa ho stvárniť slovotvorne, kompozitom alebo derivátom. Príklad. Ak v sieti



vyjadríme ATRIBUTÍV s RELÁTORom cez LVAL a PARTITÍV cez RVAL, dostaneme: *krásne dievča so zlatými vlasmi*. Ak na vzťah RELÁTORa použijeme kompozičnú funkciu CVAL a ATRIBUTÍV s PARTITÍVOM vyjadríme LVALom, máme: *krásne zlatovlasé dievča*. Ak aj partitív vyjadríme derivačne (cez DVAL), vznikne: *krásna zlatovláska*. Ak ale vyjadríme ATRIBUTÍV derivátom, PARTITÍV cez LVAL a RELÁTOR kompozitom, dostaneme: *zlatovlasá kráska*.

12.4 Príklady transformácií

ATR/PAR/REL

dievč-	LRR	krásne dievča s vlasmi ako zlato
ATR» krás-	LRL	krásne dievča so zlatými vlasmi
«PAR vlas- +pl	LLC	krásne zlatovlasé dievča
REL» zlat-	LDC	krásna zlatovláska
	DLC	zlatovlasá kráska
	DRL	kráska so zlatými vlasmi

ATR/PDK

hra-	VV	zvonky smutne hrajú
ATR» smutn-	LR	smutná hra zvonkov
PDK» zvonk-	LL	smutná zvonková hra
	LC	smutná zvonkohra

RC zvonkohra smútku
CL zvonková smutnohra
RR *hra smútku zvonkov

PCS/AGS

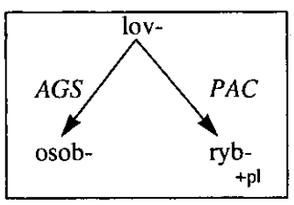
Ivan-
«PCS st-$^{+pret}$
«AGS jes-$^{+pret}$

VV Ivan stál a jedol.
VV Ivan stojac jedol.
LV Stojaci Ivan jedol.
VV Ivan pri jedle stál.
VV Ivan jedol postojačky.
VV Ivan jedol stojmo.
VV Ivan jedol v stoji.
VV Ivan jedol na stojáka.

miest-
«LOK$_1$ osob-$^{+pl}$
OKU» látk-$^{+pl}$
«PDK voň-$^{+incl}$

LOK/OKU/PDK/+INCL

DDDD voňavkárstvo
RDDD miesto, kde sú voňavkári
RRDD miesto, kde sú osoby, čo sa zaoberajú voňavkami
RRLD miesto, kde sú osoby, čo sa zaoberajú voňavými látkami
RRRR miesto kde sú osoby, čo sa zaoberajú látkami, čo majú sklon voňať



AGS/PAC

	[domin: lov]	*[domin: osoba]*	*[domin: ryby]*
VV	osoba loví ryby		
VL			osobou lovené ryby
VR			ryby lovené osobou
			ryby, ktoré loví osoba
VC	osoba *rybolovi*		
VD			

	[domin: lov]	*[domin: osoba]*	*[domin: ryby]*
LV	osobne loví ryby	ryby loviaca osoba	
LL	osobný rybí lov		osobne lovené ryby
LR	osobný lov rýb osobný lov na ryby		ryby, ktoré sú osobne lovené
LC	osobný rybolov	ryboloviaca osoba	
LD	osobná rybačka	**rybolovná osoba*	
RV		osoba loviaca ryby osoba, ktorá loví ryby	
RL	rybí lov osoby	osoba na rybom love	
RR	lov rýb osobou	osoba na love rýb	
RC	rybolov osoby	osoba na rybolove osoba, ktorá rybolovi	
DD	rybačka osoby	osoba na rybačke	
CV			
CL		rybný lovec	
CR		lovec rýb	
CC			
CD		rybolovec,*rybolovník*	
DV			
DL			**osobolovené ryby*
DR	**osobolovenie rýb*		
DC		**lovryba*	
DD			

12.5 Vznik dvojznačných parafráz

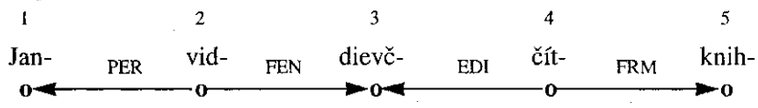
Počas syntézy SAPFO nedokáže kontrolovať, či nevznikajú viacznačné parafrázy. Nasledovné parafrázy vznikli deagentizáciou. Veta *Knihy sa kupovali* je v poriadku, ale pri *Kulaci sa prenasledovali* nie je jasné, či ich niekto prenasledoval, alebo sa prenasledovali sami:

Ľudia kupovali knihy.		Ľudia prenasledovali kulakov.	
Knihy sa kupovali	Dea	Kulaci sa prenasledovali	Dea

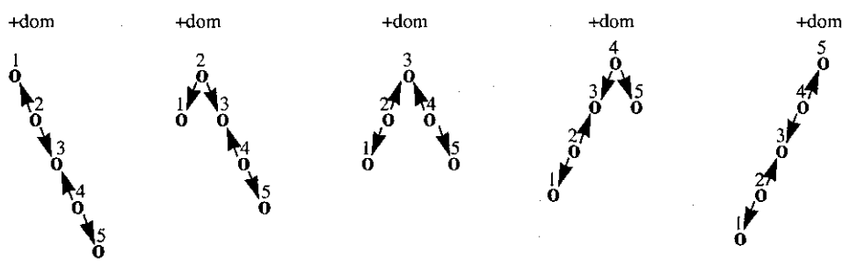
12.6 Transdominácia

Každá veta (fráza) má svoj syntakticky dominantný (centrálny) člen. Zmenou dominantného člena vznikajú trans*dominované* parafrázy.

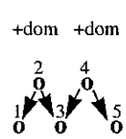
Ak žiaden z vrcholov v sieti nie je označený príznakom *+dom*, môžeme pôvodnú sémantickú sieť chápať ako vertikálne nerozčlenenú a reprezentovať ju vodorovne (alebo plošne):



Za dominantný vetný člen potom možno zvoliť ľubovoľný vrchol siete. Prvý, druhý, tretí, štvrtý alebo piaty. V prvom prípade ním bude *Jano*, v druhom jeho *videnie*, v treťom videné *dievča*, v štvrtom jej *čítanie* a v piatom ňou čítaná *kniha*. Zvyšné vetné členy sa už dominantnému členu podraďujú jednoznačne: členy so vzdialenosťou od centrálneho vrchola = *n* sa ocitnú v *n*-tej syntaktickej úrovni vnorenia:



Podľa toho, či sa rozhodneme pre centrálny člen slovesný alebo nominálny, vznikne nám buď veta alebo nominálna fráza. Dominantných členov možno zvoliť aj viacero. Napríklad druhý a štvrtý. Dostaneme dve samostatné vety s centrálnymi slovesami *vidieť* a *čítať*:



DOM	PER/FEN/EDI/FRM
1	RVRV Jano, ktorý vidí dievča, ktoré číta knihu.
1	RVLV Jano, ktorý vidí knihu čítajúce dievča.
2	VVRV Jano vidí dievča čítajúce knihu.
2	VVRV Jano vidí dievča, ktoré číta knihu.
2	VVRV Jano vidí dievča ako číta knihu.

2	VVLV	Jano vidí knihu čítajúce dievča.
2	VVDR	Jano vidí čitateľku knihy.
2	LRRV	Janovo videnie dievčaťa čítajúceho knihu.
2	LRRV	Janov pohľad na dievča, ktoré číta knihu.
3	VRLV	Knihu čítajúce dievča videné Janom.
3	VRRV	Dievča, ktoré číta knihu a ktoré Jano vidí.
4	VRVV	Dievča, ktoré Jano vidí, číta knihu.
4	RRRR	?Čítanie knihy dievčaťom videným Janom.
5	VRVR	Kniha čítaná dievčaťom, ktorú Jano vidí.
5	VLVR	Kniha, ktorú Janom videné dievča číta.
2+4	VVVV	Dievča číta knihu. Jano ju vidí.

12.7 Konverzia

Medzi slovesami existujú tzv. konverzné dvojice (Ondrejovič, 1989). Konverznými dvojicami sú napríklad *kúpiť - predať, vzať si - vydať sa, darovať - dostať, prehrať - vyhrať, učiť - učiť sa, vyvážať - dovážať, premeniť sa - vzniknúť, tvoriť - skladať sa*. Takéto dvojice dejov sa vyznačujú tým, že sa vzájomne implikujú. Ak *si Janko vzal Marienku*, tak *Marienka sa vydala za Janka;* ak *Peter predal Pavlovi auto*, tak *Pavol kúpil auto od Petra* a pod. Ide o dva spôsoby vyjadrenia jedného obsahu, teda konverzné vety treba považovať za parafrázy a každý dialógový a parafrázovací systém by mal nájsť spôsob, ako identifikovať súvislosť medzi nimi.

Jedným z riešení by bolo zaviesť inferenčné pravidlo, ktoré transformuje AGENS prvého deja na BENEFAKTOR druhého deja, a predložkový pád *od* + *gen* na AGENS druhého deja:

X kúpil Y od Z-a ---► Z predal Y X-ovi

Navrhujeme ponechať si techniku inferenčných pravidiel až na dedukciu a indukciu *skutočne nových* poznatkov nad databankou a riešiť úlohu o konverzii elegantnejším spôsobom:

Vychádzajúc z princípu jedno-jednoznačného priradenia medzi reprezentáciou a obsahom, sa na konverzné dvojice nebudeme dívať ako na dva rôzne deje, ale len ako na dve slovesá vyjadrujúce jeden a ten istý dej, pretože to jeden dej aj *fakticky* je. Tento dej je raz vyjadrený z pohľadu darcu, inokedy prijímateľa. V role aktívneho AGENSa zobrazujeme v prvom prípade darcu, v druhom prijímateľa.

V tomto zmysle nebudeme mať v sieti dva vrcholy pre *kúpiť* a *predať*, ale iba jeden vrchol pre dej *kúpiť-predať*. Zadefinujeme aj štvoricu hĺbkových pádov, ktorá je spoločným menovateľom oboch dejov: DONOR (darca), RECIPIENT (prijímateľ), OBJEKT (premiestňovaný majetok), ANTIOBJEKT (protihodnota). A napíšeme intenčné rámce oboch slovies:

kúpiť				**predať**			
0	+ nom	=	REC	0	+ nom	=	DON
0	+ ak	=	OBJ	0	+ ak	=	OBJ
od	+ gen	=	DON	0	+ dat	=	REC
za	+ ak	=	ABJ	za	+ ak	=	ABJ
na	+ ak	=	MOT	na	+ ak	=	MOT

Zašifrujme vetu *Peter kúpil od Pavla auto* podľa rámca pre *kúpiť*. Dostaneme doluuvedenú SRL-reprezentáciu. Ak teraz použijeme pravidlá v rámci pre *predať* sprava doľava a SRL-sieť dešifrujeme, vyjde nám povrchovo úplne iná veta: *Pavol predal Petrovi auto*:

kúp-pred-$^{+pret\ +perf}$
DON» Pavol-
REC» Peter-
OBJ» aut-

Peter kúpil auto od Pavla.
Pavol predal auto Petrovi.

Uvedená sieť teda reprezentuje dve veľmi odlišné, ale významovo blízke vety naraz. Rámce pre *kúpiť* a *predať* sú ako dva pohľady na jednu skutočnosť - podľa toho, cez ktorý rámec sa na sieť dívame, dostávame rôzne povrchové vyjadrenia tohto faktu.

V snahe previesť čo najviac inteligencie na samotnú reprezentáciu, budeme pokračovať aj ďalej.

J. Panevová (1978) upozorňuje, že veta *Peter kúpil od Pavla auto s nadšením* sa po konverzii nemení na vetu *Pavol predal Petrovi auto s nadšením*, t. j. že sémantické participanty po konverzii nezostávajú na svojom mieste.

Avšak vetný člen *s nadšením* vôbec medzi slovesné participanty nepatrí! V kapitole 5.36 sme ukázali, že deverbatívne substantívum v predložkovom páde *s + inštrumentál* je jednou z foriem vyjadrenia doplnku a na najvrchnejšej úrovni organizácie vety stojí iba syntakticky. Sémantický analyzátor SAPFO ho zaradí ako druhú predikáciu k subjektu alebo k objektu, takže reprezentačná sieť bude vyzerať takto:

kupov-predáv-$^{+pret\ +perf}$
DON» Pavol-
REC» Peter-
«AFE nadchýn-$^{+pret\ +perf}$
OBJ» aut-

Peter kúpil od Pavla auto. Peter bol nadšený.
Pavol predal Petrovi auto. Peter bol nadšený.

Konverziou sa konvertujú aktanty na prvej úrovni vnorenia ako celky. Pritom druhá predikácia zostáva v oboch prípadoch u *Petra*.

12.8 Perspektivizácia

So sémantickými pádmi sme doteraz pracovali ako s hĺbkovými pádmi. Najpodstatnejšou vlastnosťou *hĺbkového* pádu je, že aj po zmene povrchovej formy zostáva rovnaký.

Videli sme však, že konverziou sa sémantické pády menia. Bolo treba zvoliť osobitnú množinu *skutočne hĺbkových* pádov, ktoré boli vzhľadom na všetky povrchové transpozície vrátane konverzie invariantné.

Podobným spôsobom zasahuje sémantickú štruktúru vety aj perspektivizácia:

Napchať fajku tabakom.	*Včely sa hemžia v záhrade.*
Napchať tabak do fajky.	*Záhrada sa hemží včelami.*
Zomlieť zrno na múku.	*Pekár vysypal múku.*
Namlieť zo zrna múku.	*Múka sa pekárovi vysypala.*
Tancovať celú noc.	*Deti stratili pred otcom úctu.*
Pretancovať celú noc.	*Otec stratil u detí úctu.*

Fajku chápeme raz ako PACIENS a *tabak* ako REALIZÁTOR *napchatia*, inokedy je PACIENSom *tabak* a *fajka* LOKATÍVom *napchatia*. *Pekár* je zobrazený raz ako úmyselný AGENS *vysypania*, druhý raz ako neúmyselný BENEFAKTOR.

Pripomeňme si, že kritériom funkčnosti celého nášho systému je jeho schopnosť parafrázovať. Avšak vzhľadom na toto kritérium sú potom sémantické pády volené nefunkčne, pretože parafrázovanie neumožňujú. V prácach zo sémantickej syntaxe (Miko, 1962; Oravec, 1967; Daneš, 1981; Mluvnice češtiny, 1987) sa pracuje s vetnými vzorcami a gramatické pády sa uvádzajú do vzťahu k určitým sémantickým pádom. Tieto sú však pre účely parafrázovania nefunkčné: Ak *piecť chlieb* a *piecť múku* je v oboch prípadoch zásahový PACIENS, alebo v *kosiť trávu na lúke* je *tráva* PACIENS a *lúka* OKOLNOSŤ a v *kosiť lúku s trávou* je naopak *lúka* PACIENS a *tráva* OKOLNOSŤ, ako ich nepopliesť pri parafrázovaní? Ak vo vetách

Tancovala celú noc
Pretancovala celú noc

je *noc* raz PACIENSom a druhý raz iba OKOLNOSŤou, nič nenaznačuje, že medzi oboma vetami je akákoľvek súvislosť. Neumožňuje to uvedomiť si *noc* z prvej vety a *noc* z druhej vety ako totožné.

Ak nasledujúce štyri vety referujú o jednej udalosti

Okradli suseda.
Ukradli peniaze.
Vykradli suseda.
Vykradli obchod.

a v každej má ísť o PACIENS, tak PACIENS nie je hĺbkovým pádom. Ak vo vete

Zlodeji ukradli susedovi peniaze z obchodu

sú *peniaze* OBJEKT, *sused* BENEFAKTOR a *obchod* LOKATÍV, potom ak má ísť o hĺbkové pády, participanty musia zostať v týchto rolách aj po konverzii. No pri takejto voľbe pádov po konverzii nastane zmätok a dostávame vety ako *Zlodeji okradli peniaze o suseda* a pod.

12.9 Faktické pády

Jediná možnosť je pokladať používané sémantické pády za akúsi významovú medzirovinu (Panevová, 1986) a zaviesť ešte jednu (obsahovú) rovinu so skutočne hĺbkovými pádmi, ktoré stoja v jedno-jednoznačnom priradení ku skutočnostiam a sú vzhľadom na konverziu a perspektivizáciu invariantné.

Tieto pády nazveme *faktické pády* a na rozdiel od sémantických pádov ich budeme označovať tučným písmom. Tým sa však úloha porozumieť vzťahom v jazyku presúva do novej roviny, ktorá je ešte panenskou pôdou: treba vypracovať akúsi *faktickú syntax*, ktorá definuje vzťahy gramatických pádov k pádom faktickým; resp. sémantická syntax definuje vzťahy gramatických pádov k sémantickým a faktická syntax by mala preskúmať vzťah sémantických pádov k faktickým.

Ak chceme zachytiť povedzme fungovanie slovesa *kradnúť* nielen samého osebe, ale aspoň v kontexte jeho odvodenín *ukradnúť, okradnúť, vykradnúť,* musíme faktické pády zvoliť odlišne od sémantických a valenčné rovnice definovať napríklad takto:

ukradnúť (306)

VAL(306,0,nom,živ,AGS,**E1**).
VAL(306,0,ak,kon,OBJ,**E2**).
VAL(306,0,dat,živ,BEN,**E3**).
VAL(306,z,gen,lok,LOK,**E4**).

vykradnúť (308)

VAL(308,0,nom,živ,AGS,**E1**).
VAL(308,0,ak,lok,OBJ,**E4**).
VAL(308,0,ak,živ,OBJ,**E3**).
VAL(308,0,dat,živ,BEN,**E3**).

okradnúť (307)

VAL(307,0,nom,živ,AGS,**E1**).
VAL(307,0,ak,živ,OBJ,**E3**).
VAL(307,o,ak,kon,REA,**E2**).

Na deji sa zúčastňujú vždy štyri entity **E1**, **E2**, **E3**, **E4**. Pritom **E1** je vždy tá, ktorá čin spácha, **E2** vždy to, čo je premiestňované, **E3** je pôvodný majiteľ a **E4** miesto, odkiaľ **E2** odcudzia. Potom ak chceme externalizovať štruktúru

(vy)(u)(o)kradn-
E1» zlodej-$^{+pl}$
E2» peniaz-$^{+pl}$
E3» sused-
E4» obchod-

môžeme sa podľa potreby rozhodnúť, či sa chceme dívať na **E3** ako na BENEFAKTOR alebo PACIENS, prípadne ktorú z entít **E2**, **E3**, **E4** zdôrazníme tým, že jej vyhradíme OBJEKTovú rolu. Podľa toho potom zvolíme príslušné sloveso.

Aby konverzné rovnice boli úplné, zavedieme ešte inferenčné pravidlo:

-kradn- -kradn-
E3» X ---------► **E4**» Y
E4» Y POS» X

To nám umožní transponovať **E3** do roly POSESORa. Jedenásť valenčných rovníc s inferenčným pravidlom potom generuje tieto parafrázy:

Zlodeji ukradli susedovi peniaze z obchodu.
Zlodeji okradli suseda o peniaze.
Zlodeji vykradli suseda.
Zlodeji vykradli susedovi obchod.
Zlodeji vykradli susedov obchod s peniazmi.
Zlodeji ukradli peniaze zo susedovho obchodu.

Faktický pád má teda označovať faktický vzťah medzi dvoma entitami ako taký, v celej jeho plnosti. Preto sa často nedá ani pomenovať. Pomenovanie znamená už jazykové, sémantické stvárnenie faktu vyzdvihnutím niektorého z jeho čiastkových aspektov. Napríklad faktický vzťah vo fráze *vodné jazero* by sme mohli označiť len **LOK-MAT**, t. j. LOKATÍV a MATERIÁL zároveň. Na sémantickej úrovni si potom môžeme a musíme vybrať a vysloviť iba jeden z nich: buď prezentujeme *vodu* ako MATERIÁL *jazera*, alebo *jazero* ako LOKATÍV *vody*. Podobnými faktickými pádmi by mohli byť aj **OTEC-DCÉRA, BRAT-SESTRA**, ktoré sa dajú sémanticky stvárniť vždy iba z jednej svojej stránky.

12.10 Ambivalencia doplnku

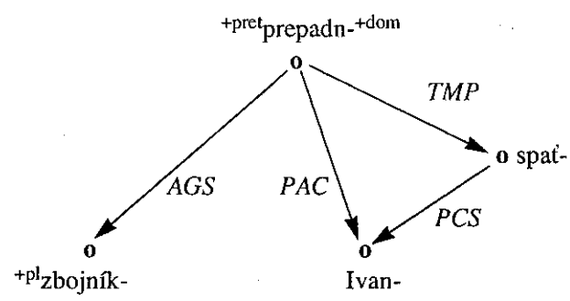
Až doteraz sme zvykli vety reprezentovať syntaktickými a sémantickými stromami (t. j. acyklickými grafmi). V strome je každý z členov podradený práve jed-

nému nadradenému členu. Avšak *doplnok* je vetný člen, ktorý môže mať *dva* nadradené členy. Tým narúša stromovú reprezentáciu a donucuje, aby sme reprezentačný aparát rozšírili na grafy ako také, vrátane cyklických sietí:

„Doplnok je vetný člen, ktorý sa vzťahuje dvoma rozličnými vzťahmi na dva vetné členy zároveň, a to na podmet a prísudok, alebo na predmet a prísudok" (Kačala, 1971).

Napríklad vo vete *Zbojníci prepadli Ivana v spánku* možno *spánok/spať* chápať ako druhotnú predikáciu, činnosť Ivana, ale aj ako časovú okolnosť *prepadnutia*. Ak vetu nechceme ochudobniť o časť jej významu, musíme reprezentovať oba vzťahy naraz (TEMPORÁL aj PROCESOR). Vznikne cyklická sieť:

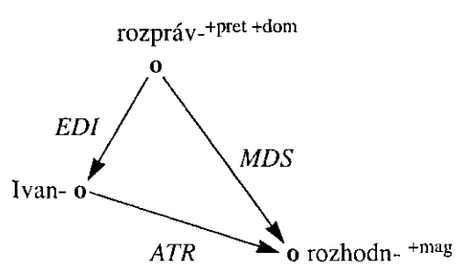
Zbojníci prepadli Ivana v spánku.



Zbojníci prepadli Ivana. Ivan spal.
Zbojníci prepadli Ivana keď spal.

Vo vete *Ivan rozprával s veľkou rozhodnosťou* sa prívlastok *rozhodnosti* vzťahuje tak na *Ivana* ako aj na jeho *rozprávanie*; a vetu možno oboma spôsobmi aj parafrázovať:

Ivan rozprával s veľkou rozhodnosťou.



Ivan rozprával. Ivan bol veľmi rozhodný.
Ivan rozprával rozhodne.
Ivanovo rozprávanie bolo veľmi rozhodné.

12.11 Selekcia

Pretože syntaktický syntezátor SAPFO dokáže prijímať a externalizovať (meniť na povrchové štruktúry) iba acyklické stromy (inak by sa algoritmus zacyklil), musí každému generovaniu viet predchádzať selekcia hrán. *Selekcia* spočíva v tom, že zo sémantickej siete povynechávame hrany ľubovoľným takým spôsobom, aby sa zmenila z cyklického grafu na acyklický. Z každého cyklu vynecháme povedzme jednu hranu.

Z posledne uvedenej siete možno vynechať buď hranu *ATR*, pričom dostaneme parafrázu *Ivan rozpráva veľmi rozhodne*, alebo vynecháme *MDS* a výsledkom bude parafráza *Ivan rozpráva a je veľmi rozhodný*. Hranu *EDI* môžeme vynechať tiež, ale žiadna parafráza nevznikne, lebo *rozprávaniu* potom chýba obligátny subjekt.

Zo zložitejšej siete možno hrán vynechať aj viac. Dostaneme nesúvislý graf s viacerými acyklickými komponentmi, ktorých počet sa bude rovnať počtu vygenerovaných samostatných viet.

12.12 Komplexné vetné členy

Skúmajme štvoricu parafráz:

Jano vidí ***padať biely sneh.***
Jano vidí biely ***sneh*** *ako padá.*
Jano vidí ***padanie*** *bieleho snehu.*
Jano vidí ***belosť*** *padajúceho snehu.*

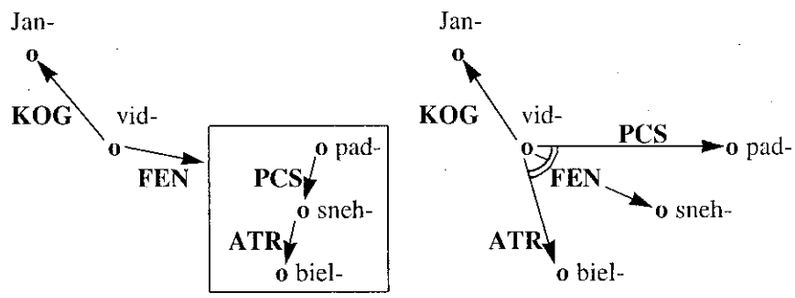
Vidíme, že vety sú obsahovými parafrázami jedna druhej. Čo je však v skutočnosti objektom *Janovho videnia* v prvej vete? Posledné tri vety sú vzájomne transdominované. Ale ktorá z nich je bezprostrednou trans*formačnou* parafrázou prvej vety? Je to *sneh*, jeho *padanie* alebo jeho *biela farba*, ktorú Jano vidí?

Jano vidí *sneh*, jeho *padanie* aj jeho *belobu*, pretože *sneh* nemožno vnímať oddelene od jeho dynamických a statických príznakov. Objekt vnímania fakticky splýva, a preto nie je bezpodmienečne nevyhnutné diferencovať ho ani sémanticky. Žiaden poslucháč nepožiada o spresnenie, či teda Jano vidí sneh alebo jeho padanie.

Padanie snehu je jediný *komplexný* faktický objekt, ktorý v prvej z viet splýva aj sémanticky. Komplexné vetné členy budeme reprezentovať tak, že sa faktický pád, v ktorom stoja, bude vzťahovať na ne ako na celok:

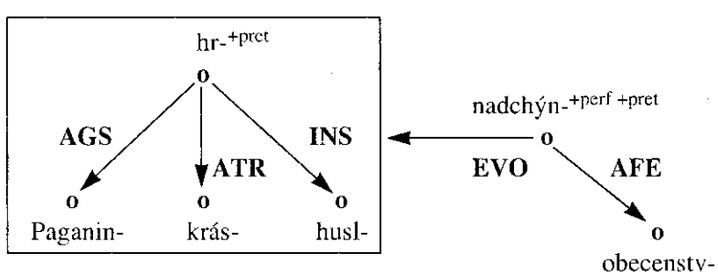
vid-
KOG» Jan-
FEN» [pad-
PCS» sneh-]
ATR» biel-

alebo:



Keby vo vete *Vidím padať sneh* bolo objektom *padanie* a vo *Vidím sestru plakať* by bola objektom *sestra*, potom by aktuálne členenie menilo hlavné vetné členy. Infinitív sa spravidla považuje za doplnok. Avšak o doplnku sme povedali, že „sa vzťahuje dvoma rozličnými vzťahmi na dva vetné členy zároveň, a to na podmet a prísudok, alebo na predmet a prísudok" (Kačala, 1971). To je inými slovami to isté, ako povedať, že prísudok sa vzťahuje jedným vzťahom na dva vetné členy zároveň. Vo vetách ako *Videl som sestru behať bosú* budú potom dva (aj viac) doplnkov, ktoré sa vzťahujú dvoma, troma a viacerými rozličnými vzťahmi na predikát a iné vetné členy zároveň. Nie je jednoduchšie povedať, že predikát sa vzťahuje *jediným* vzťahom na *trojjediný komplexný objekt: bosú bežiacu sestru?*

Komplexný môže byť aj subjekt v roli EVOKÁTORA pri slovesách sentiendi:



Paganini hral krásne na husle. Obecenstvo bolo nadšené.
Paganini nadchol obecenstvo krásnou hrou na husle.
Paganiniho husle nadchli obecenstvo krásnou hrou.
Paganiniho hra nadchla obecenstvo svojou krásou.
Krása Paganiniho hry nadchla obecenstvo.

Pravidlo 1. Faktický objekt je pri slovesách vnímania, cítenia a poznávania vždy komplexný. Členy vnútri komplexného vetného člena možno transdominovať a vzniknuté parafrázy zostanú korektné. Komplexný vetný člen je faktickým

OBJEKTOM či **KAUZÁTOROM** *ako celok*. Na sémantickej rovine môžeme a musíme zvoliť ľubovoľný, ale len jeden z jeho elementov za sémantický OBJEKT či KAUZÁTOR.

-kúzl-$^{+circ}$
EVO» [dievč-
ATR» krás-]

Krásne dievča okúzľuje.
Dievčia krása okúzľuje.
Dievča okúzľuje svojou krásou.

-hnev-
AFE» otec-
EVO» [správ- sa
PCS» syn-]

Synovo správanie hnevá otca.
Syn hnevá otca svojím správaním.

Pravidlo 2. Do komplexného faktického **OBJEKTU** sa zlieva aj sémantický OBJEKT manuálnych a iných procesov spolu s jeho LOKATÍVOM v prípade, že OBJEKT priestorovo vypĺňa celý LOKATÍV:

kos-
PAC» [tráv-
LOK» lúk-]

kosiť trávu na lúke
kosiť trávnatú lúku

uchvát-
AFE» [obecenstv-
LOK» sál-]

uchvátiť obecenstvo v sále
uchvátiť sálu s obecenstvom

pi-$^{+ex}$
PAC» [vod-
LOK$_1$» pohár-]

vypiť pohár vody
vypiť vodu v pohári

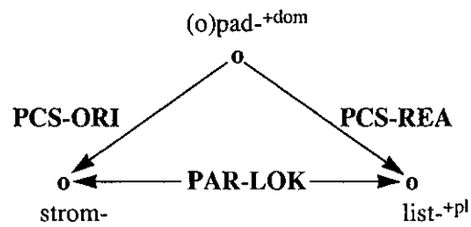
hemž- sa
PCS» [včel-$^{+pl}$
LOK$_1$» záhrad-]

Včely sa hemžia v záhrade.
Záhrada sa hemží včelami.

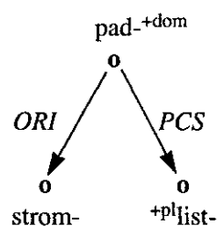
12.13 Transfigurácia

Transfiguráciou rozumieme parafrázovanie, ktoré (na rozdiel od transformácie) mení nielen formu, ale aj sémantické figúry (pády), avšak nemení faktické vzťahy. Transfigurované parafrázy vznikajú z jednej faktickej reprezentácie štiepením ambivalentných pádov, vrcholov a selekciou hranovej množiny. Typickými príkladmi transfigurácie sú konverzia a perspektivizácia.

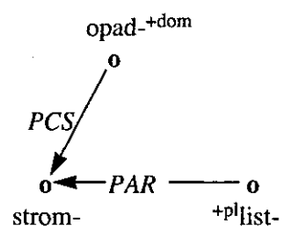
Príklad. Štiepením faktickej reprezentácie:



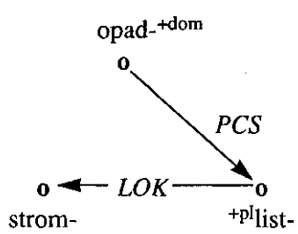
vzniknú štyri navzájom transfigurované sémantické siete:



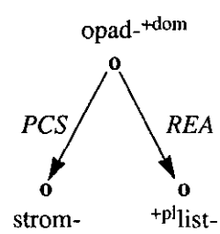
Listy padajú zo stromu.
Zo stromu padá lístie.



Listnatý strom opadáva.
Strom s lístím opadáva.



Listy stromu opadávajú.
Lístie na strome opadáva.



Strom opadáva od listov.
Strom opadáva od lístia.

12.14 Neutralizácia komponentov významu v kontexte

V hovorenej ale aj písanej reči sa často stáva, že nesynonymné lexémy a konštrukcie dočasne fungujú ako synonymné. Aj od každého dialógového systému a od systémov pre inteligentný návrat informácií z databanky sa spravidla požadujú nesynonymné odpovede na otázky.

Napríklad na výzvu

- Chcem sa naučiť po španielsky. Nepoznáš nejakého *učiteľa?*

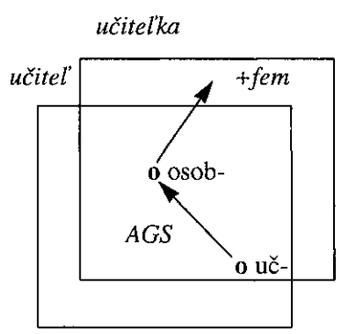
sa neodpovedá

- Žiaľ, nepoznám. Iba *učiteľku*.

ale skôr

- Pani Esperanza je dobrá *učiteľka*.

Hoci slová *učiteľ*, *učiteľka* očividne nie sú synonymá, predsa majú veľkú časť významu spoločnú. A potom časť významu, v ktorej sa líšia. Do prieniku významov patrí *'osoba, ktorá učí'*. Rozdiel významov tvorí príznak ženského rodu:



Pretože príznak ženského rodu *+fem* je vzhľadom na komunikačný zámer celého rozhovoru irelevantný, fungujú *učiteľ*, *učiteľka* v tomto kontexte ako synonymá. Inak povedané, v uvedenom rozhovore sa aktualizuje iba tá časť ich významu, ktorú majú spoločnú. Nástáva tzv. *neutralizácia sémantických diferencií v kontexte* (Apresian, 1974).

12.15 Transmutácia

Transmutáciou budeme rozumieť každé parafrázovanie, ktorého základom je eliminácia elementárnych sémantických diferencií (komponentov významu) alebo obohatenie o nové komponenty významu. Trans*mutované* parafrázy vznikajú z jednej faktickej reprezentácie eliminácioU alebo adíciou jedného alebo viacerých komponentov významu. Líšia sa teda obsahovo, avšak iba v medziach elementárnych sémantických diferencií. Obsahovú diferenciu medzi transmutovanými frázami

možno vždy vyjadriť sumou významových komponentov. Prehľad komponentov významu a príklady elementárnych transmutáciií pozri v 11.3.

12.16 Algebra syntakticko-sémanticko-faktických manipulácií s jazykom

SAPFO je teda schopné robiť súhrnne sedemdesiat druhov rozličných manipulácií so slovenským jazykom. Patria medzi ne manipulácie transmutačné, transfiguračné, transdominačné, transformačné, transfokačné a ďalšie. Osobitné postavenie majú *štylizácia, dialektizácia, frazeologizácia a básnické figúry*:

Transmutácie gramatické: *feminizácia, pluralizácia, kolektivizácia, negativizácia, kondicionalizácia, temporalizácia, aspektizácia, reflexivizácia, frekventizácia, disperzia, rezultativizácia, dynamizácia, reorientácia.*

Transmutácie lexikálno-sémantické: *synonymizácia, antonymizácia, abstrakcia, konkretizácia, magnonymizácia, minorizácia, deminutivizácia, hyperbolizácia, symbolizácia.*

Transfigurácie: *transfigurácia, perspektivizácia, konverzia, deagentizácia, depacientizácia.*

Transdominácie: *dominácia.*

Transformácie syntaktické: *pasivizácia, tranzitivizácia, infinitivizácia, participalizácia, komplementizácia, apozícia, subordinácia, koordinácia, adordinácia, solitudizácia, parentéza, reverencia.*

Transformácie kondenzačné: *ľavá nominalizácia, pravá nominalizácia, derivácia, kompozícia, juxtapozícia, univerbizácia, kontrakcia, abreviácia, akronymizácia, fonizácia, apostrofácia, numeralizácia, signácia.*

Transformácie eliptické: *koordinačná elipsa, metonymická elipsa, apoziopéza, proziopéza.*

Transfokácie: *fokácia, inverzia.*

Intonačné: *intonácia.*

Úplný zoznam a príklady syntaktických, sémantických a faktických manipulácií s jazykom sa nachádza v kapitole 13. Niektoré transmutácie sú buď priamo prevzaté, alebo sú aspoň analogické *lexikálnym parametrom* J. D. Apresiana a I. A. Meľčuka (1974).

Každá manipulácia je označená trojicou písmen. Ku každej manipulácii je okrem toho priradený zoznam komponentov významu, ktoré sa eliminujú alebo pribúdajú po jej aplikácii. Zoznam môže byť aj prázdny. Každá transmutačná parafráza mení aspoň jeden komponent významu. Transformácie a transfigurácie spravidla komponenty významu nemenia (iba v ojedinelých prípadoch). Transdominácia, transfokácia a intonácia presúvajú každá svoj špecifický komponent (príznak) - *+dom, +fok, +int.*

Sedemdesiaticu manipulácií možno chápať ako **lineárnu algebru** nad priestorom slovenských viet. *Identická transformácia* predstavuje neutrálny prvok. Ku

každej manipulácii existuje aj manipulácia *inverzná*. Manipulácie možno aplikovať na slovenské vety a frázy, pričom výsledkom aplikácie bude zasa slovenská veta (fráza). Manipulácie možno kombinovať do postupností a aplikovať rekurzívne.

12.17 Nesynonymické parafrázovanie I

P. Sgall a kol. (1986) dôsledne skúmali pojem striktnej synonymie. V našom parafrázovacom systéme sme uplatnili pojem synonymie v oveľa širšom zmysle. Keby sme vychádzali z pojmu striktnej synonymie, parafrázovacie zväzky by boli príliš nepočetné, chudobné.

Celá štruktúra systému SAPFO je daná jediným cieľom, utváraná jedinou hodnotou: úlohou SAPFO je riešiť *problém sémantickej relevancie* (14.1), t. j. vedieť rozhodnúť, kedy sú dve frázy sémanticky (a obsahovo) relevantné a kedy nie.

Videli sme, že sémanticky relevantné môžu byť v kontexte aj frázy, ktoré nemusia byť synonymné v prísnom zmysle slova. Aj transmutované frázy musíme teda považovať za parafrázy a vzťah medzi nimi chápať ako *slabú, kontextovú synonymiu*.

Medzi parafrázy musíme počítať aj negáciu a antonymizáciu, pretože ide o dvojice fráz s maximálnym prienikom významu, až na jediný komponent: znamienko záporu *+neg*. Na otázku *Akú farbu má sklo* sú odpovede *Sklo je bezfarebné*, *Sklo nemá farbu* úplne relevantné.

12.18 Definícia parafrázy

Parafrázu ľubovoľného segmentu jazyka *L* (slova, frázy, vety, súvetia, textu) definujeme ako každý taký segment jazyka *L'*, že *L* a *L'* sú sémanticky relevantné.

Sémanticky relevantné môžu byť v určitom kontexte ľubovoľné dva segmenty jazyka, ktorých významy majú neprázdny prienik.

Špeciálne v systéme SAPFO platí za parafrázu frázy *L* každá fráza *L'*, pre ktorú existuje postupnosť formálnych manipulácií $T_1, T_2, T_3, \ldots, T_n$ taká, že

$$L' = T_1 T_2 T_3 \ldots T_n L.$$

Teda množinu parafráz frázy *L* tvoria všetky frázy, ktoré vzniknú z *L* aplikáciou rôznych postupností jazykových manipulácií, a to z tých sedemdesiatich, ktoré SAPFO dokáže vykonávať.

$T_1 T_2 T_3 \ldots T_n$ potom tvorí *sémantický most* medzi *L* a *L'*.

A množina komponentov významu asociovaná s $T_1, T_2, T_3, \ldots, T_n$ tvorí *sémantickú diferenciu* medzi *L* a *L'*.

12.19 Nesynonymické parafrázovanie II

Nesynonymná definícia parafrázy však neznamená ľubovôľu alebo zmätok vo významoch. Meníme síce význam, ale zároveň si uchovávame záznam o tom, *ktorý* z významov a *ako presne* bol zmenený. Miera a smer významových posunov zostanú zakódované v postupnosti formálnych manipulácií, ktoré sme na frázu aplikovali.

Keď máme dané dve frázy *L* a *L'*, namiesto toho, aby sme rozhodovali problém, *či sú alebo nie sú synonymné*, ideme radšej skúmať, *akým spôsobom sa k sebe vzťahujú*. Hľadáme postupnosť formálno-významových manipulácií, ktoré tvoria sémantický most medzi nimi.

Potom namiesto toho, aby nám zostalo v ruke len málo účelné tvrdenie, že *L* a *L'* sú alebo nie sú synonymné (čo nevysvetľuje o nich v podstate nič pozitívne), môžeme porozumieť vzťahu medzi *L* a *L'* v termínoch elementárnych manipulácií formy a významu. Postupnosť $T_1T_2T_3...T_n$ tvorí most (cestu) od *L* k *L'*, rozčlenenú pekne analyticky a prehľadne na elementárne, jedno-krokové posuny významu a formy. A množina k nim asociovaných komponentov významu predstavuje sémantický rozdiel *L'- L*.

Myslíme si, že ak chceme lepšie porozumieť procesu jazykovej komunikácie ako celku, treba venovať viac pozornosti práve nesynonymickým vzťahom v jazyku.

12.20 Príklady nesynonymných parafráz

fráza a parafrázy	*sémantický most*	*sémantická diferencia*
prší		
neprší	^Neg	*L' = -L*
popŕcha	^Min	*L' = L +min*
leje	^Mag	*L' = L +mag*
pŕška	^NomMin	*L' = L +min*
lejak	^NomMag	*L' = L +mag*
bude pršať	^Tem	*L' = L +fut*
zaprší	^Tem	*L' = L +fut*
pršalo	^Tem	*L' = L +pret*
zapršalo	^TemAsp	*L' = L +pret +perf*
pršiavalo	^TemFre	*L' = L +pret +iter*
rozpršalo sa	^TemFaz	*L' = L +pret +inc*
zlatý prsteň		
zlatistý prsteň	^NolMin	*L' = L +min*
zlatý prstienok	^NolDem	*L' = L +min*
?zlatný prsteň	^NolSyn	*L' ≈ L*
?prsteň do zlatova	^Nor	*L' = L +min*

zlaté dievča		
milé dievča	^NolSyn	$L' \approx L$
milá dievčica	^NolSynStl	$L' \approx L$ *+arch*
milá deva	^NolSynStl	$L' \approx L$ *+poet*
zlatulinké dievčatko	^NolDemMag	$L' = L$ *+min +mag*
červený drak		
karmínový drak	^NolSyn	$L' \approx L$
červená dračica	^NolFem	$L' = L$ *+fem*
ohnivočervený drak	^NolMag	$L' = L$ *+mag*
drak červený ako oheň	^NorMag	$L' = L$ *+mag*
?draci s červeňou	^NorPlu	$L' = L$ *+plu*
ryba		
rybka	^DerMin	$L' = L$ *+min*
rybička	^DerMin^2	$L' = L$ *+min*^{2}
?veľrybička	^Der^{2}Min^2	$L' = L$ *+min*^{2} *+mag*
Africká flóra		
rastlinstvo Afriky	^NorSyn	$L' \approx L$
stromy v Afrike	^NorKol^{-1}Abs^{-1}	$L' \leq L$
?africký smrek	^NolKol^{-1}Abs^{-2}Plu^{-1}	$L' \leq L$ *-plu*
mladý človek		
starec	^DerAnt	$L' = -L$
mládež	^DerKol	$L' = L$ *+plu*
mladší človek	^NolGra	$L' = L$ *+maj*
nie najstarší človek	^NolLit	$L' = L$ *+min*
letieť		
lietať	^Der	$L' = L$ *+dir*
lietavať	^Der	$L' = L$ *+iter*
dolietať	^Der^2	$L' = L$ *+dir +fin +perf*
doletieť	^Der	$L' = L$ *+fin +perf*
zaletieť	^Der	$L' = L$ *+post +perf*
uletieť	^Der	$L' = L$ *+ab +perf*
odletieť	^Der	$L' = L$ *+ab +perf*
odlietať	^Der^2	$L' = L$ *+ab*
odlietavať	^Der^2	$L' = L$ *+ab +dir*
odletovať	^Der^2	$L' = L$ *+ab +iter*
obletieť	^Der	$L' = L$ *+circ +perf*
obletovať	^Der^2	$L' = L$ *+circ +iter*
vyletieť	^Der	$L' = L$ *+ex +perf*
vylietať	^Der^2	$L' = L$ *+ex*

preletieť	^Der	*L' = L +trans +perf*
prelietať	^Der^2	*L' = L +trans*
prelietať	^Der^2	*L' = L +trans +dir +perf*
prelietavať	^Der^2	*L' = L +trans +dir*
preletovať	^Der^2	*L' = L +trans +iter*
popreletovať	^Der^3	*L' = L +dis +trans +iter*
priletieť	^Der	*L' = L +ad +perf*
podletieť	^Der	*L'= L +sub +perf*
vletieť	^Der	*L' = L +in +perf*
nadletieť	^Der	*L' = L +supra +perf*
naletieť	^Der	*L' = L +sur +perf*
nalietať	^Der^2	*L' = L +rez +perf*
nalietať sa	^Der^2	*L' = L +int +perf*
vzlietnuť	^Der	*L' = L +alt +perf*
zletieť	^Der	*L' = L +deo +perf*
zletieť sa	^Der^2	*L' = L +con +perf*
rozletieť sa	^Der^2	*L' = L +dsp +perf*
porozletovať sa	^Der^3	*L' = L +dis +dsp +perf*

12.21 Rozdiel medzi transformáciou, transfiguráciou a transmutáciou

Transformácia znamená presuny na povrchovej (syntaktickej) rovine. *Transfigurácia* znamená presuny na významovej (sémantickej) rovine. A *transmutácia* presuny na obsahovej (faktickej) rovine.

Transformácia mení gramatické pády (formu) a sémantické necháva nezmenené. Transfigurácia mení zúčastnené sémantické pády (figúry) a faktické necháva nezmenené. Transmutácia mení aj faktické pády.

Transformácia mení spôsob, *ako* niečo hovoríme, pričom to, čo hovoríme, zostáva rovnaké. Transfigurácia ponecháva to, *o čom* hovoríme, ale mení to, *čo o tom* hovoríme. Transmutácia pozmeňuje (mutuje) samotný predmet rozhovoru.

12.22 Kovanie čistého významu do foriem

Obsah prechádza siedmimi fundamentálnymi procesmi, než sa stane formou:

obsah	**1. Selekcia**	*o čom* chceme hovoriť
	2. Mutácia	
význam	**3. Figurácia**	*čo* o tom chceme povedať
	4. Dominácia	
	5. Formácia	
forma	**6. Fokácia**	*ako* to chceme povedať
	7. Intonácia	

Na počiatku je úplný systém poznatkov. Konkrétne napríklad uzavreté univerzum poznatkov z nejakého vedného odvetvia alebo ľubovoľný logicky ucelený príbeh. Databanka poznatkov pozostáva z jednej komplexnej siete, ktorá reprezentuje sumu zúčastnených entít a faktických vzťahov medzi nimi. Predstavuje pôvodnú plnosť informácie, jedinosť, z ktorej sa štiepením a postupným ochudobňovaním vytvárajú miliardy čiastkových vyjadrovacích foriem.

Selekcia. Nech pôvodná sieť pozostáva z *m* vrcholov (entít). Z nich vyberieme *n* entít (vrcholov), o ktorých chceme hovoriť. To môžeme urobiť $\binom{m}{n}$ spôsobmi. Dostaneme podsieť s *n* vrcholmi.

Nech v tejto n-prvkovej podsieti existuje *k* hrán (vzťahov). Z nich vyberieme *l* vzťahov (hrán), o ktorých budeme hovoriť. To môžeme urobiť najviac $\binom{k}{l}$ spôsobmi. Dostaneme podgraf s *n* vrcholmi a *l* hranami.

Hrany siete eliminujeme tak, aby sme dostali *acyklický* súvislý podgraf (strom) alebo niekoľko acyklických súvislých komponentov (les). Počet súvislých komponentov grafu sa bude rovnať počtu viet, ktoré nakoniec dostaneme. Ak podgraf pozostáva z jediného komponentu (stromu), výsledkom bude práve jedna veta. Ak je to les, výpoveď bude pozostávať z toľkých viet, koľko je v lese stromov. (Terminológiu z teórie grafov pozri v Znám (1980).

Celková parafrázovacia potencia selekcie teda činí:

$$\mathbf{P}_{SEL} \leq \sum_{i=1}^{m} \binom{m}{i} \cdot \sum_{j=1}^{k} \binom{k}{j} = 2^m 2^k$$

Nie každá selekcia sa bude môcť úspešne sformovať do súvetia. Napríklad niektoré syntaktické formy explicitne vyžadujú vyjadriť určité obligatórne aktanty. Ak sme ich počas selekcie vynechali, alebo vzťahy vedúce k nim, vyjadrovací proces zlyhá na úrovni formácie.

Mutácia. Počas mutácie sa kladie základ pre všetko *ne*synonymické parafrázovanie. Zo sietí pozmenených (mutovaných) v tomto štádiu vzniknú výpovede, ktoré sa budú odlišovať v obsahu.

Ku každému vrcholu siete sú pridružené komponenty významu. Jednotlivé komponenty možno eliminovať, alebo pripojiť nové. Okrem toho samotný vrchol možno zmeniť v rámci synonymie, antonymie, hyperonymie, hyponymie a negativizáciou. Mutovať možno jeden alebo viac vrcholov a pridružených komponentov.

Zmenu *i* prvkov budeme nazývať mutáciou *i*-teho rádu. Keďže SAPFO pracuje s cca 66-timi komponentmi významu, celková parafrázovacia potencia mutácie *i*-teho rádu bude činiť:

$$\mathbf{P}_{MUT} \leq (66.n)^i$$

Skutočný počet možných mutácií bude oveľa menší. Je ohraničený počtom synoným, antoným, hyperoným, hyponým a derivátov prítomných v slovníkoch a kompatibilitou komponentov významu s konkrétnymi lexiami. Napríklad príznak

deminutívnosti sa môže aktualizovať iba pri koreňoch, pri ktorých sa v slovotvornom hniezde deminutíva naozaj nachádzajú; ku koreňom vyjadrujúcim nejaký dej nemá význam pripájať príznak ženskosti *+fem*; iné korene sú zasa nezlučiteľné s príznakmi smerovania deja *+ex, +in, +ab* a pod. Takto mutované siete nenájdu na úrovni formácie formu, ktorou by sa vyjadrili.

Voľbou komponentov významu na tejto úrovni možno ovplyvniť a upraviť vzhľad súvetia v zmysle niektorého štýlu alebo dialektu.

Figurácia. Počas figurácie sa utvárajú základné sémantické figúry. Holý fakt dostáva sémantické stvárnenie, hovoriaci doň vnáša svoj vlastný postoj, dôraz, perspektívu. Z totality faktických vzťahov vyberá iba tie, ktoré sú pre neho dôležité, ktoré chce podčiarknuť.

Pri figurácii sa z faktických pádov na hranách odštiepia sémantické pády. Pri obsažnejších faktických pádoch je to možné viacerými spôsobmi.

Z vrcholov s konverznými, dvoj- a viacperspektívnymi dejmi a vlastnosťami sa odštiepi jedna z perspektív, vo svetle ktorej budeme dej (vlastnosť) prezentovať.

Na tejto úrovni transfigurované vety sa budú líšiť sémantickou perspektívou. Počet možných transfigurácií je ohraničený počtom viacperspektívnych hrán a vrcholov:

$$P_{FIG} \leq 3^{nl}$$

Dominácia. Ľubovoľný z *n* vrcholov môžeme zvoliť za dominantný. Ten sa potom stane centrom budúcej vety alebo nominálnej frázy a ostatné sa mu jednoznačným spôsobom podradia. Vrcholy, ktoré majú od centrálneho vrcholu vzdialenosť = 1 sa stanú hlavnými aktantmi vety, resp. prívlastkami mennej frázy na prvej, najvyššej úrovni syntaktickej organizácie vety. Entity so vzdialenosťou od centrálneho vrchola rovnou *i* sa zúčastnia vo vetnej stavbe na *i*-tej úrovni vnorenia.

N-vrcholový graf (strom), ktorý vznikol počas selekcie, sa počas dominácie usporiada vertikálne. Vzťahy podradenosti a nadradenosti vetných členov sa na tejto úrovni fixujú. Horizontálne sa vetné členy zoradia až pri fokácii.

Transdominované vety sa líšia centrálnym vetným členom. Ľahko nahliadneme, že počet transdominácií sa rovná počtu vetných členov:

$$P_{DOM} = n$$

Formácia. Na tejto úrovni sa význam začína kryštalizovať do konkrétnych foriem.

Amorfné jadrá slovotvorných hniezd sa konštituujú v rámci niektorého zo slovných druhov ako konkrétne slová. Zo sémantických pádov sa odštiepia pády syntaktické a gramatické. Slová nadobudnú určitý morfologický tvar.

Predovšetkým treba rozhodnúť, či centrálny člen má byť slovesný alebo nominálny. Podľa toho výpoveď nadobudne formu vety alebo mennej frázy.

Formáciu kontrolujú kolokačné valenčné funkcie VAL, LVAL, RVAL, CVAL, DVAL. Každá z nich dokáže dať sémantickému participantu viacero foriem. VAL

dokáže sformovať aktanty do tvaru mennej frázy v siedmich pádoch, do predložkovej frázy, príslovky, infinitívnej konštrukcie a vedľajšej vety. Cez LVAL prechádzajú všetky zhodné prívlastky: adjektíva, číslovky, príčastia a zhodné doplnky. RVAL dáva participantom formu mennej, predložkovej alebo adjektívnej frázy, prístavku, príslovky, infinitívnej alebo prechodníkovej konštrukcie, vedľajšej vety a parentézy. Ak zveríme nejaký participant funkciám CVAL, DVAL, pokúsia sa pridružiť ho k nadradenému členu slovotvorne. Ak je to možné, spravidla existuje jeden spôsob, ako to urobiť pomocou derivátu, a jeden či dva spôsoby pomocou kompozita. To dáva spolu 27 foriem.

Transformované vety (frázy) sa líšia svojou slovesnosťou (mennosťou) a rozdielnym výberom syntaktických vyjadrovacích prostriedkov. Pre počet tranformácií *n*-člennej vety platí horný odhad:

$$\mathbf{P}_{FOR} \leq 2.27^n$$

Skutočný počet prípustných transformácií má k hornému odhadu ďaleko. Jedna voľba syntaktických prostriedkov podmieňuje a obmedzuje druhú; jednotlivé voľby teda nie sú nezávislé. V závislosti od toho, či zvolíme pre nadradený člen mennú alebo slovesnú formu, môžeme bezprostredne mu podradené členy vyjadriť len ako aktanty, resp. prívlastky. A ak sú niektoré syntaktické pozície už obsadené, iné participanty sa prostredníctvom nich už nemôžu realizovať.

Fokácia. Poradím vetných členov vyjadrujeme aktuálne členenie vety na tému a ohnisko. Zmena slovosledu je možná aj vnútri nominálnych fráz. N-prvkový strom, ktorý sme vybrali pri selekcii, a ktorý dostal svoje vertikálne usporiadanie počas dominácie, sa na tejto úrovni usporadúva horizontálne. Transfokované vety sa líšia slovosledom.

Parafrázovacia potencia fokácie je daná počtom permutácií aktantov vo vete a prívlastkov vnútri nominálnych fráz. V krajnom prípade to môže byť:

$$\mathbf{P}_{FOK} \leq n!$$

Intonácia. Predpokladajme najjednoduchší intonačný model: jeden z vetných členov môže niesť zvukový dôraz, vyznačený kurzívou. Potom:

$$\mathbf{P}_{INT} = n$$

12.23 Analógia s metalurgiou

Proces premeny obsahu na formu sa podobá odlievaniu, kutiu a opracovávaniu kovov.

Selekcia znamená výber materiálu, s ktorým hodláme pracovať. Mutáciou namiešame zliatinu v správnom pomere a môžeme pridať stopové komponenty (legovanie), ktoré ovplyvnia výsledné vlastnosti materiálu. Figurácia zodpovedá

odliatiu do formy. Po figurácii možno žeravý kov ohýbať, ale v zásade je dané, *čo* bude výsledným produktom. Ohýbaniu mäkkého kovu zodpovedá dominácia. Formácia znázorňuje kutie a úplné ochladenie kovu, ktorý tak nadobudne definitívny tvar. Nakoniec fokácia zodpovedá brúseniu a opracovávaniu kovu za studena a intonácia povrchovej úprave, lešteniu, natieraniu.

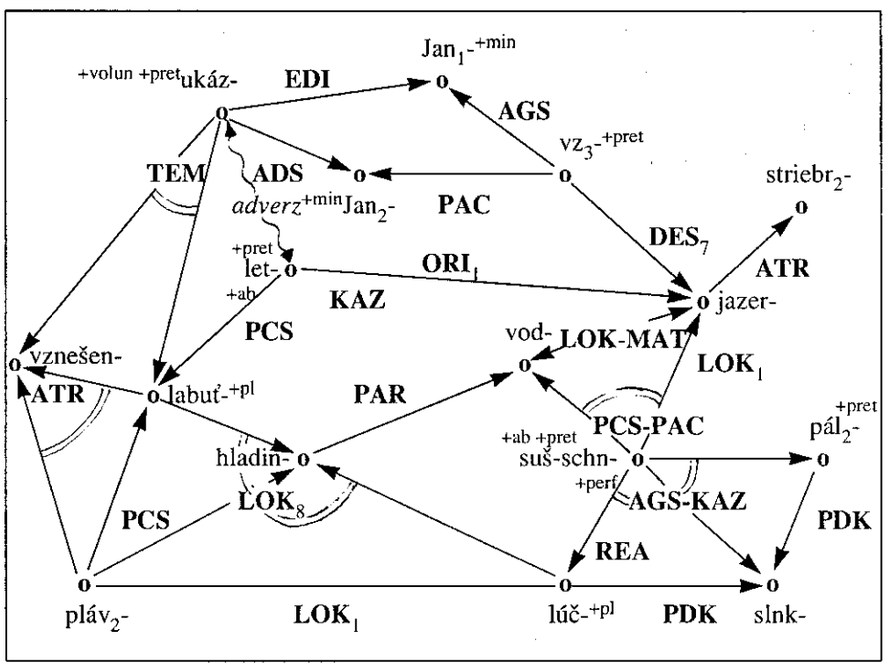
Tekuté skupenstvo zodpovedá plnosti cítenia, ktorú nemôžeme vysloviť. Je tehotné množstvom foriem. Tuhé skupenstvo zodpovedá vyjadreniu, ktoré je však vždy už iba čiastočné.

12.24 Veľký príklad

Celý proces premeny obsahu na formu - tak, ako sa odohráva v našom počítačovom modeli - si ukážeme na príklade.

Vezmime za základ tento krátky príbeh:

[1]



Janko vzal Janku k striebornému jazeru. Chcel jej ukázať labute, s akou vznešenosťou plávajú v slnečných lúčoch na vodnej hladine. Ale labute už odleteli. Slnko pálilo a voda v jazere vyschla.

Sieť sa vyznačuje istým stupňom informačnej redundancie; mohla vzniknúť interpretáciou obrazov alebo navrstvením výpovedí viacerých subjektov, z ktorých každý vnímal udalosť svojím spôsobom.

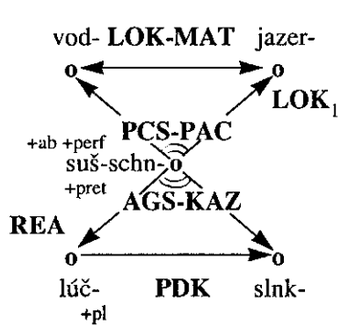
Všimnite si, že viacerým faktickým pádom a entitám je ponechaná ich ambivalencia (viacznačnosť), čo - ako uvidíme - nebude na prekážku, ale pri parafrázovaní nám to padne na úžitok:

TEMATÍV ukázania:	*ukázať vznešenosť labutí*
	ukázať vznešené labute
PACIENS vysušenia:	*vysušiť vodu*
	vysušiť jazero
AGENS vysušenia:	*slnečná páľava vysušila vodu*
	slnečné lúče vysušili vodu
	slnko vysušilo vodu
hladina ako **LOKATÍV**$_8$:	*labute na hladine*
	plávanie na hladine
	lúče na hladine
vznešenosť ako **ATRIBÚT**:	*vznešenosť labutí*
	vznešenosť plávania

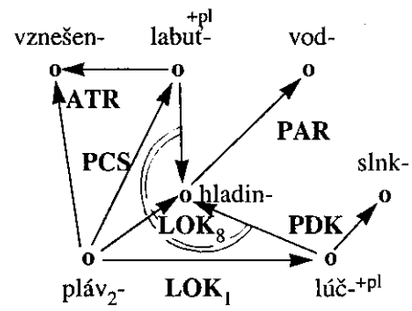
Faktický pád medzi *vodou* a *jazerom* je označený ako **LOK-MAT**, pretože *jazero* je miestom, kde sa *voda* nachádza, a *voda* je materiálom, z ktorého *jazero* pozostáva. Medzi *sušením* a *vodou* je faktická rola **PCS-PAC**, pretože *voda* môže vzhľadom na dvojperspektívny konverzný dej *sušiť-schnúť* vystupovať v sémantickej role PROCESORa *vysychania* alebo PACIENSa *vysúšania*. Vzťah medzi *sušením-schnutím* a jazerom nesie označenie **LOK-PCS-PAC**, pretože *jazero* je miestom, kde sa odohráva *vysychanie-vysušovanie vody*, ale zároveň aj subjektom *vysychania* a objektom *vysušovania*. A faktický pád medzi *lúčmi, pálením, slnkom* a *sušením-schnutím* má memotechnickú podobu **AGS-KAZ**, pretože *slnko*, jeho *pálenie* a *lúče* možno sémanticky stvárniť buď ako AGENS *vysušovania*, alebo ako KAUZÁTOR *vysychania*. Vzťah medzi *lúčmi* a *sušením-schnutím* je navyše rozšírený na **AGS-KAZ-REA**, pretože *lúče* možno chápať buď ako primárny AGENS *vysušovania*, alebo len ako sekundárnu substanciu, prostredníctvom ktorej REALIZUJE *vysušovanie slnko*.

Selekcia. Vyberme zo siete **[1]** podgraf **[1.1]** pozostávajúci z piatich vrcholov: *vod-jazer-suš/schn-lúč-slnk*. Iným výberom by mohol vzniknúť napríklad **[1.2]**:

[1.1]

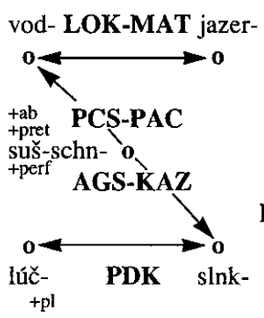


[1.2]

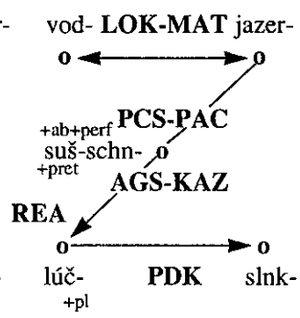


Zo siete [1.1] môžeme povynechávať hrany tak, aby sme dostali acyklický podgraf s jedným alebo viacerými komponentmi (strom alebo les):

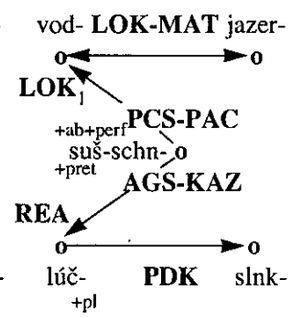
[1.1.1]



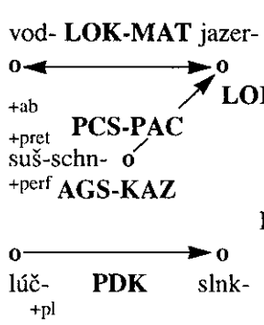
[1.1.2]



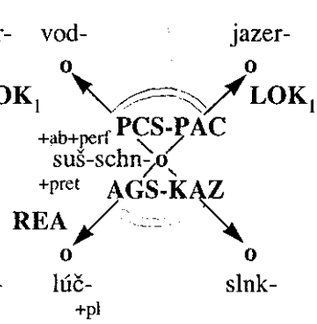
[1.1.3]



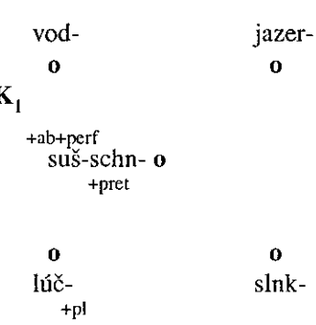
[1.1.4]



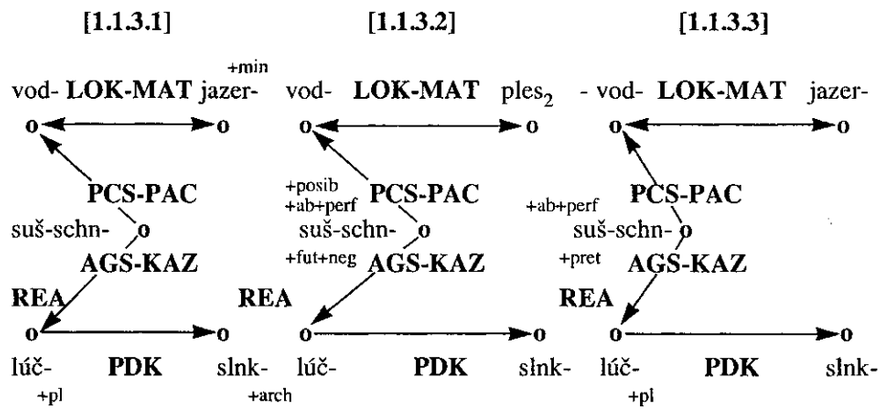
[1.1.5]



[1.1.6]

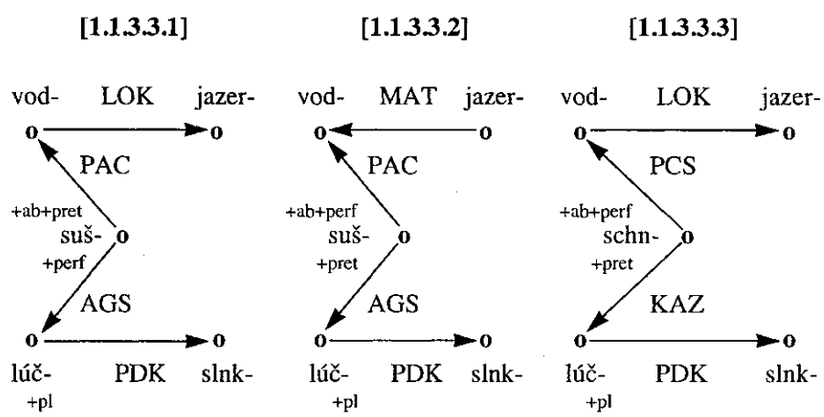


Mutácia. Vyberme sieť [1.1.3]. Môžeme ju mutovať najrôznejšími spôsobmi:



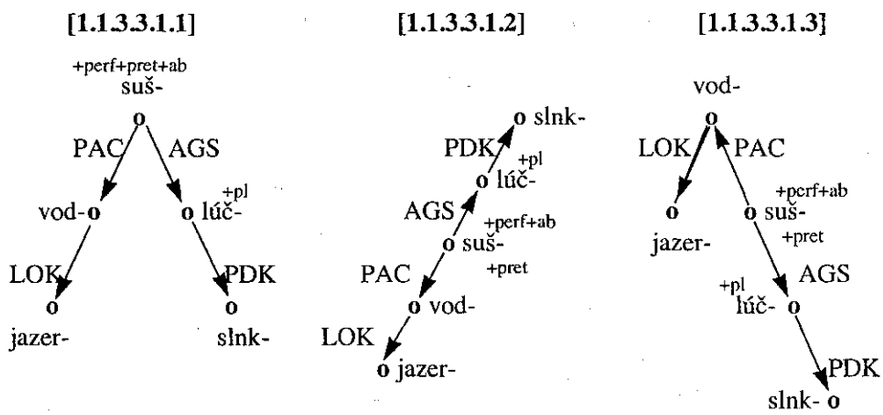
Pokračujme s nezmenenou sieťou [1.1.3.3].

Figurácia. Počas figurácie sa faktické pády a entity štiepia na jednu zo svojich stránok:



Zvoľme si napríklad alternatívu [1.1.3.3.1]: Dej zobrazíme v tranzitívnej perspektíve (*sušiť koho čo*), pričom *lúče* budú hrať sémantickú rolu AGENSa, *voda* bude PACIENS a *jazero* LOKATÍV *vody*.

Dominácia. Dominantný člen môžeme zvoliť piatimi spôsobmi. Sieť tak získa vertikálne usporiadanie:



Zvoľme alternatívu **[1.1.3.3.1.1]**. Centrálnym členom bude *suš-*.

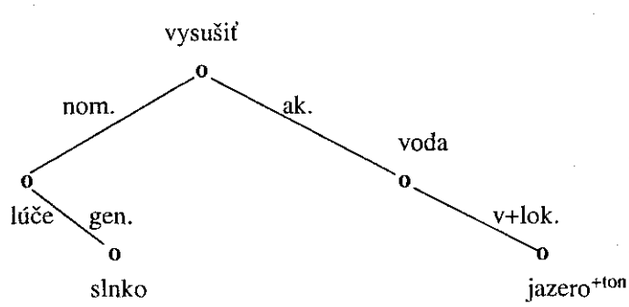
Formácia. Nech sa sieť **[1.1.3.3.1.1]** stane vetou. Vrchol *suš-*$^{+perf+pret}$ sa podľa toho musí sformovať ako dokonavé určité sloveso v minulom čase. V slovotvornom hniezde *suš-* sa nájde sloveso *vysušiť*.

Funkcia VAL dovoľuje vyjadriť AGENS nominatívom a PACIENS akuzatívom alebo pasívne inštrumentálom a nominatívom. LOKATÍV a PRODUKTOR sa ďalej musia premeniť na prívlastky hlavných členov. Podľa nominalizačných vzorcov sa *slnk-* môže prejaviť ako vzťahové adjektívum *slnečný*, holý genitív *slnka* alebo predložkový pád *z* + genitív. *Jazer-* buď v tvare *jazerný*, alebo genitívom či predložkovým pádom *v* + lokál.

Zvoľme holý genitív pre *slnko* a v+lokál pre *jazero*. Vetu si želáme mať v aktívnom tvare.

Fokácia a intonácia. Nech *lúče* sú témou a *voda* rémou vety. Na začiatku vety bude teda stáť AGENS. PACIENS bude v ohnisku na konci vety. Touto voľbou syntaktický strom nadobúda horizontálne usporiadanie:

[1.1.3.3.1.1.1]



Intonačnými prostriedkami zdôrazníme, že ide o vodu v *jazere*, nie inde. Konečne dostávame vetu:

Lúče slnka vysušili vodu ***v jazere.***

Intonačne by sme mohli zdôrazniť aj ktorýkoľvek iný člen:

Lúče ***slnka*** *vysušili vodu v jazere.*
Lúče slnka vysušili ***vodu*** *v jazere.*

Po transfokácii AGENSa do ohniska by veta znela:

Vodu v jazere vysušili lúče slnka.

Počas formácie sme ďalej mali na výber medzi mennou a slovesnou formou, pasívnou a aktívnou perspektívou a tiež medzi viacerými prívlastkovými formami. Transformačný zväzok siete **[1.1.3.3.1.1]** by mal podobu:

Vysušenie vody v jazere slnečnými lúčmi.
Vysušenie jazernej vody lúčmi slnka.
Lúče zo slnka vysušili vodu jazera.
Slnečné lúče vysušili jazernú vodu.
Voda v jazere bola vysušená slnečnými lúčmi.

Ak by sme v štádiu dominácie boli zvolili iný dominantný člen, vznikali by transdominované frázy (prvé štyri sú zo siete **[1.1.3.3.1.3]**):

Jazerná voda vysušená lúčmi slnka.
Voda jazera, ktorú vysušili slnečné lúče.
Jazerná voda, ktorá bola vysušená lúčmi zo slnka.
Slnečnými lúčmi vysušená voda jazera.
Jazero, ktorého vodu vysušili lúče slnka.
Jazero, kde slnečné lúče vysušili vodu.
Slnečné lúče, čo vysušili jazernú vodu.
Lúče zo slnka, ktorými bola vysušená voda v jazere.
Slnko, ktorého lúče vysušili vodu v jazere.

Ak sa vrátime na úroveň figurácie, dostávame transfigurované parafrázy. Sieť **[1.1.3.3.3]** prezentuje *vodu* ako reflexívny PROCESOR, pričom *slnečné lúče* skĺznu do roly KAUZATÍVu:

Voda v jazere vyschla kvôli slnečným lúčom.
Lúče slnka spôsobili vyschnutie jazernej vody.
Voda jazera sa vysušila pod lúčmi slnka.

Ak sa vrátime až na úroveň mutácie, môžeme pokračovať rozvíjaním sietí **[1.1.3.1]** a **[1.1.3.2]**. Eliminujeme ablatív a dokonavosť slovesa, zmeníme čas, štýl a pod. Transmutované parafrázy sa začnú rozchádzať obsahovo:

Lúče slnca vysušili vodu v jazierku.
Slnečné lúče nebudú môcť vysušiť jazernú vodu.
Slnečné lúče usušili jazernú vodu.
Jazerná voda vysychala pod lúčmi zo slnka.

Odlišnou počiatočnou selekciou hrán vzniká ešte širšia rodina parafráz. Z [**1.1.1**] a [**1.1.2**]:

Slnko usušilo vodu v jazere.
Voda jazera sa sušila na slnku.
Slnečné lúče vysušili vodné jazero.
Kvôli lúčom slnka vyschlo jazero vody.
Jazero s vodou schne pod lúčmi zo slnka.

[**1.1.4**] pozostáva z dvoch komponentov. Z jedného vznikne veta, z druhého menná fráza:

Lúče slnka. Vodné jazero vyschlo.
Slnečné lúče. Jazero s vodou sa suší.

V [**1.1.5**] sme vynechali asi dosť podstatné slúvislosti:

Slnko a lúče vysušia jazero a vodu.
Jazero s vodou vysychajú kvôli slnku a lúčom.

Krajným prípadom je sieť [**1.1.6**], z ktorej sme vynechali všetky hrany. Vzťahy nebudú vyjadrené:

Slnko. Lúče. Vysušenie. Jazero. Voda.

Aj túto päticu fráz považujeme za plauzibilnú parafrázu siete [**1.1**]. Mohol by ju vysloviť napríklad niekto, kto by od vzrušenia nad vyschnutím vody nebol schopný súvislého prejavu.

Konečne transselekciou vrcholov môžeme zmeniť tému rozhovoru. Počet transselekcií je daný počtom rozličných volieb vrcholovej podmnožiny a počtom všetkých možných ciest po hranách medzi nimi. Napríklad parafrázovaním [**1.2**] dostaneme:

Vznešene plávajú labute v slnečných lúčoch na vodnej hladine.
Vznešené labute plávajú na hladine vody v lúčoch slnka.
Labute plávajú v slnečných lúčoch s vznešenosťou.
Vznešenosť labutí na hladine vody.

12.25 Strojová poézia

Nasledovné dve dvojice parafráz sa rýmujú. Vznikli *transformačnými symetriami* (cyklickou zámenou poradia aplikácie valenčných funkcií). SAPFO tým dokázala, že si zaslúži svoje meno, ktoré dostala na počesť prvej známej ženy v dejinách literatúry:

smutná zvonkohra	^KpzNol
zvonková smutnohra	^NolKpz
krásna zlatovláska	^KpzDerNol
zlatovlasá kráska	^KpzNolDer

13.

PREHĽAD SYNTAKTICKÝCH, SÉMANTICKÝCH A FAKTICKÝCH MANIPULÁCIÍ S JAZYKOM

Transmutácie

Gramatické transmutácie

Feminizácia *Fem* [+fem]

herec
herečka Fem

krava
býk Fem^{-1}

Pluralizácia *Plu* [+pl]

človek
ľudia Plu

Kolektivizácia *Kol* [+pl]

list
lístie Kol

admirál
admiralita Kol

rastlina
rastlinstvo Kol
flóra Kol

Negativizácia *Neg* [-X]

závislý
nezávislý Neg

farebný	
bezfarebný	Neg
pokoj	
nepokoj	Neg
rozumiem	
nerozumiem	Neg

Kondicionalizácia *Knd* [+kond,des]

robím	
robil by som	Knd

Temporalizácia *Tmp* [+pret,+fut]

kvitne	
kvitol	Tmp
pokvitne	Tmp
bude kvitnúť	Tmp

Aspektizácia *Asp* [+perf]

robiť	
urobiť	Asp

Frekventizácia *Fre* [+iter]

chodím	
chodievam	Fre

Disperzia *Dsp* [+dsp]

letím	
lietam	Dsp

Reflexivizácia *Ref* [+ref]

blázniť	
blázniť sa	Ref

Rezultativizácia *Rez* [+rez]

pustý hrad	
spustnutý hrad	Rez

Dynamizácia *Dyn* [+dyn]

jama je čierna	
jama sa černie	Dyn

Reorientácia *Reo* [+ex,in,ad,ab,...]

Jano išiel.	
Jano odišiel.	Reo
Jano vyšiel.	Reo
Jano prišiel.	Reo
Jano vošiel.	Reo

Lexikálno-sémantické transmutácie

Lexikálno-sémantické transformácie sa vykonávajú hlavne zámenou lexií za pomoci lexikálno-sémantických parametrov *dublet (X,Y), synon$_i$(X,Y), anton$_i$(X,Y), abstr(X,Y), symbol(X,Y).* Deminutivizáciu a minorizáciu vykonáva derivačná funkcia DVAL, magnonymizácia sa deje sčasti na základe slovníkového lexikálneho parametra *magnon$_i$(X,Y)*, sčasti slovotvorne.

Hoci vzťah medzi synonymami v slovníku nie je špecifikovaný, niekedy sa vykryštalizuje súčinnostne. Ak napríklad štylistický slovník identifikuje jedno zo synoným ako štylisticky alebo dialektologicky príznačné, ide o štylistické, resp. nárečové synonymá a sémantická diferencia medzi nimi je daná sémantickým komponentom štylistickej/nárečovej príznačnosti:

vesna = jar +poet
sľifka = slivka +orava

Ak ide o obsahové synonymá, nie sme schopní sémantickú diferenciu algoritmicky určiť, a po transformácii píšeme už len znamienko približnej rovnosti:

stolička ≈ taburetka

Identifikovať rozdiel medzi *stoličkou* a *taburetkou* ako *neprítomnosť operadla* sa možno pokúsiť na základe konotačnej databanky, ktorá môže obsahovať aj celý

vysvetľovací slovník slovenského jazyka. To je však výpočtovo extrémne náročné, výsledok je nespoľahlivý a závisí od momentálneho obsahu databanky.

Abstrakcia rozširuje rozsah pojmu a mení sémantickú rovnosť na nerovnosť:

ruža ≤ kvet

Aj v tomto prípade je sémantická diferencia nevypočitateľná. Po zreťazení abstrakcie a konkretizácie sa nerovnosti menia na približnú rovnosť:

ruža ≈ ľalia

Synonymizácia *Syn* [≈],[+expr,arch,...]

priezračný	
priesvitný	Syn
priehľadný	Syn
ísť po sekeru	
ísť pre sekeru	Syn

Antonymizácia *Ant* [-X]

drahý prsteň	
lacný prsteň	Ant
nad stolom	
pod stolom	Ant

Abstrakcia *Abs* [≥]

kráčať	
pohybovať sa	Abs
mahagónový	
drevený	Abs
sokol	
vták	Abs
dravec	Abs^2
zviera	Abs^3

pred dverami
pri dverách Abs

Konkretizácia *Abs*$^{-1}$ [≤]

kvet
ruža Abs^{-1}

Magnonymizácia *Mag* [+mag]

strom
stromisko Mag

ticho
mŕtve ticho Mag

biely
snehobiely Mag

veľký
velikánsky Mag
veličizný Mag

svätý
svätosvätý Mag

prší
leje Mag

Minorizácia *Min* [+min]

kniha
knižka Min
knižočka Min^{2}

sladký
sladkastý Min
nasladlý Min

spievať
pospevovať si Min

Deminutivizácia *Dem* [+cit]

kniha	
knižka	Dem
knižočka	Dem^2

pekný	
peknučký	Dem
peknulinký	Dem^2

spievať	
spievkať	Dem

Symbolizácia *Sym* [≈]

dar z lásky	
dar zo srdca	Sym

Peter je ľstivý	
Peter je líška	Sym

Hyperbolizácia *Hyp* [+mag]

malý	
malililinký	Hyp

dlhý	
dlhááááansky	Hyp

Transfigurácie

Vlastná transfigurácia *Tfg*

Paganini hral krásne na husle. Obecenstvo v sále bolo uchvátené.	
Paganini uchvátil obecenstvo v sále krásnou hrou na husle.	Tfg
Paganiniho husle uchvátili obecenstvo v sále krásnou hrou.	Tfg
Paganiniho hra uchvátila obecenstvo v sále svojou krásou.	Tfg
Krása Paganiniho hry uchvátila obecenstvo v sále.	Tfg
Paganini uchvátil sálu s obecenstvom krásnou hrou na husle.	Tfg^2

Slnečné lúče pálili a jazero sa vysušilo.
Pálenie slnečných lúčov vysušilo jazero. Tfg
Slnečné lúče vysušili jazero. Tfg
Slnko vysušilo jazero. Tfg

Jano poranil Fera ostrým kameňom.
Janov kameň poranil Fera svojou ostrosťou. Tfg
Ostrosť Janovho kameňa poranila Fera. Tfg

kosiť trávnatú lúku
kosiť trávu na lúke Tfg

Listy stromu opadávajú.
Zo stromu padá lístie. Tfg
Listnatý strom opadáva. Tfg

Perspektivizácia *Per* [-itc,...]

Listy absorbujú slnečné svetlo.
V listoch sa absorbuje slnečné svetlo. Per

Natrieť chlieb maslom.
Natrieť maslo na chlieb. Per

Chlapca zaujíma umenie.
Chlapec sa zaujíma o umenie. Per

Učiteľ má u žiakov rešpekt.
Žiaci majú pred učiteľom rešpekt. Per

Zlodeji ukradli susedovi peniaze.
Zlodeji okradli suseda o peniaze. Per

Katka ponúkla Petrovi víno.
Katka ponúkla Petra vínom. Per

Pekár vysypal múku z vreca.
Pekárovi sa vysypala múka z vreca. Per

Peter požičal Pavlovi knihu.
Pavol si požičal knihu od Petra. Per

Konverzia *Kon*

Peter predal Pavlovi auto.
Pavol kúpil od Petra auto. Kon

Janko si vzal Marienku.
Marienka sa vydala za Janka. Kon

Bojí sa tmy.
Tma ho ľaká. Kon

Obraz je nad stolom.
Pod obrazom je stôl. Kon

Karol je vyšší od Petra.
Peter je nižší od Karola. Kon

Farby tvoria spektrum.
Spektrum sa skladá z farieb. Kon

Deagentizácia *Dea*

Deti zjedli čerešne.
Čerešne sa zjedli. Dea

Depacientizácia *Dep*

Rodičia stavali dom ťažko.
Rodičom sa stavalo ťažko. Dep

Transdominácia

Dominácia *Dom*

prísny otec
otcova prísnosť DomNol
prísnosť otca DomNor

siedmi trpaslíci
sedmica trpaslíkov DomNor

splynený koks
koksový plyn DomNol

zuhoľnatené drevo
drevené uhlie DomNol

Transformácie

Syntaktické transformácie

Pasivizácia *Pas*

Voda zaplavila lúku.
Lúka je zaplavená vodou. Pas

Tranzitivizácia *Tra* [+sim,+lib]

Anna tancuje a spieva.
Anna tancuje spievajúc. Tra

Infinitivizácia *Inf*

Prídem, aby som sa pozreľ.
Prídem sa pozrieť. Inf

Participalizácia *Par* [sim]

Jano vidí dievča. Dievča číta knihu.
Jano vidí dievča čítajúce knihu. Par

Komplementizácia *Kpl*

Sestra sa vrátila domov. Sestra je tichá.
Sestra sa vrátila domov tichá. Kpl

Apozícia *Apo*

Ganga je rieka. Ganga je modrá.
Rieka Ganga je modrá. Apo

Subordinácia *Sub*

Učiteľ rozprával.
učiteľ, ktorý rozprával Sub

Koordinácia *Kor*

Anna tancuje. Anna spieva.
Anna tancuje a spieva. Kor

Adordinácia *Ado*

Pod lesom bolo jazierko. Pod lesom bol rybník.
Pod lesom bolo jazierko, rybník. Ado

Solitudizácia *Sol* [+expr]

Retardácia, osamostatnenie vetného člena. Zvyšuje dôraz. Čiastočne sa kryje s doplnkom.

Dívali sa na seba cez slzy.
Dívali sa na seba - cez slzy. Sol
Dívali sa na seba. Cez slzy. Sol

Parentéza *Ptz*

O indukcii budeme ďalej hovoriť podrobne v kapitole o magnetizme.
O indukcii budeme ďalej hovoriť podrobne (v kapitole o magnetizme). Ptz

Reverencia *Rev* [+rev]

Vykanie a onikanie.

Prosím ťa, nerob to!
Prosím vás, nerobte to! Rev
Prosím ich, nerobia to! Rev

Kondenzačné transformácie

Ľavá nominalizácia *Nol*

Ján prichádza
Jánov príchod Nol

Pravá nominalizácia *Nor*

Ján prichádza
príchod Jána Nor

Derivácia *Der*

zmeniť sa na kameň
skamenieť Der

putujúci človek
pútnik Der

miesto, kde je zvon
zvonica Der

polovičný mesiac
polmesiac Der

Kompozícia *Kpz*

pekné písanie
krasopis Kpz

dievča so zlatými vlasmi
zlatovlasé dievča Kpz

častice β
β-častice Kpz

Juxtapozícia *Jux*

Sprahovanie (juxtapozícia) sa od skladania (kompozície) líši tým, že nejde o spojenie slovných koreňov, ale celých gramatických tvarov, ktoré si pritom zachovávajú aj svoje gramatické kategórie (pád, číslo, zápor) a sú iba juxtaponované (t. j. umiestnené vedľa seba bez medzery). Počet juxtaponovaných členov môže byť dva alebo aj viac.

Pri analýze spriahnutého tvaru stačí identifikovať švíky, čiže hranice medzi pôvodnými slovami. Po doplnení medzerníkov vzniká plnohodnotná syntaktická konštrukcia, ktorú možno bez všetkého odovzdať syntaktickému analyzátoru.

jeden a dvadsať
jedenadvadsať Jux

vstanie z mŕtvych
zmŕtvychvstanie Jux

neschopný práce
práceneschopný Jux

od včera
odvčera Jux

Univerbizácia *Uni* [+colq]

študent priemyslovej školy
priemyslovák Uni

ulica februárového víťazstva
februárka Uni

Kontrakcia *Ktr* [+colq]

motoristický hotel
motel Ktr

zelenina-ovocie
zelovoc Ktr

matematika
matika Ktr

kilogram
kilo Ktr

Alexandra
Sandra Ktr

firmy
fy Ktr

Abreviácia *Abr* [+admin,sci]

pán Novák
p. Novák Abr

vedúci oddelenia
ved. odd. Abr[2]

československej armády
čs. armády Abr

Akronymizácia *Akr* [+admin,sci]

Vysoká škola múzických umení
VŠMU Akr

Fonizácia *Fon* [+colq]

ZNB
esenbé Fon

Apostrofácia *Asf* [+poet]

keby si
kebys' Asf

Numeralizácia *Num* [+admin,sci]

12 bratov
dvanásť bratov Num[-1]
tucet bratov Num[-1]

Signácia *Sig* [+sci]

20 percent
20 % Sig

1 plus 1
1 + 1 Sig

50 Wattov
50 W Sig

Eliptické transformácie

Elipsa *Eli*

Janko má rád jahodovú zmrzlinu a Anka má rada vanilkovú zmrzlinu.
Janko má rád jahodovú zmrzlinu a Anka má rada vanilkovú. Eli
Janko má rád jahodovú zmrzlinu a Anka vanilkovú. Eli[2]

Metonymická elipsa *Met* [≈]

Dokončil som písanie knihy
Dokončil som knihu. Met

Apoziopéza *Aza* [+cit]

Apoziopéza a proziopéza sú typické pre ústne prejavy ale aj pre ich literárne spracovanie. Vznikajú jednoducho odtrhnutím začiatku alebo konca vety. Pri apoziopéze je algoritmicky nerozhodnuteľné, či je spôsobená citovým zaujatím alebo vonkajším zásahom.

Ale raz sa všetko na dobré obráti.
Ale raz... Aza

Proziopéza *Pza*

To je moja manželka.
...moja manželka. Pza

Kontaminácia *Knt*

Chápeš to?
Chápeš tomu? Knt

povedať človeku
povedať človekovi Knt

Vzájomnou kontamináciou dvoch foriem vzniká defektná syntax. Takáto kontaminácia je okazionálna v ústnych prejavoch, ale systematická u detí a cudzincov, ktorým zatiaľ unikajú menej frekventované jazykové vzory. Táto transformácia je zaujímavá hlavne pri tvorbe počítačových systémov pre korekciu gramatiky (gram-

mar checker, grammar tutor). Takéto systémy mávajú zabudovanú zložku, ktorá predvída najčastejšie pisárske a kompozičné chyby alebo chyby typické pre cudzincov.

Transfokácia

Fokácia *Fok*

Ohnisko vety možno špecifikovať podľa desaťčlenného menu. Voľba potom ovplyvní jej aktuálne vetné členenie:		(1) agens (2) adresát (3) paciens (4) inštrument (5) čas
Jano dal Marienke kvety.	fok = pac	(6) miesto
Jano dal kvety Marienke.	fok = ads	(7) okolnosť
Marienke dal kvety Jano.	fok = ags	(8) spôsob (9) miera
Hodinky sa našli.	fok = akt	(0) regulárny
Našli sa hodinky.	fok = pac	slovosled

Inverzia *Inv* [+rog,cit,sci]

Inverzia slovosledu je možná aj vnútri nominálnych fráz. Vo všeobecnosti spĺňa rozličné funkcie. Inverzia podmetu s prísudkom je príznakom interogatívnej modálnosti, inverzia determinatívnych prívlastkov v nominálnej fráze je príznačná pre odborný štýl, inverzia hodnotiacich prívlastkoch znamená citovú emfázu.

Jano príde.	
Príde Jano?	Inv
kremičitý oxid	
oxid kremičitý	Inv
moje zlaté dievčatko	
dievčatko moje zlaté	Inv

Štylizácia

Štylizácia *Stl*

Štýl je voliteľný podľa sedemčlenného menu. Lexémy, gramatické formy a syntak-	(1) hovorový (2) literárny

tické konštrukcie nesú štylistické príznaky. Ich voľba ovplyvňuje vzhľad parafrázy:

(3) technický
(4) poetický
(5) archaický
(6) administratívny
(0) nešpecifikovaný

dievča		
dievka	Syn	[+arch]
dievčina	Syn	[+arch]
deva	Syn	[+poet]
dievčica	Syn	[+poet]

jesť		
žrať	Syn	[+pej]
papať	Syn	[+cit]

prísť do Parku kultúry a oddychu		
dostaviť sa do PKO	StlAkr	[+admin]
skočiť do pekáča	StlAkrFon	[+slang]

Dialektizácia *Dia*

dobrí synovia		
dobre sinove	Dia	[+zemplín]
dobrí sinovā	Dia	[+orava]
dobrí sinové	Dia	[+trnava]

Frazeologizácia *Fra* [≈,+expr,...]

Má hlad.	
Cigáni mu v bruchu vyhrávajú.	Fra

Ivanovi to prekáža.	
Ivanovi to je tŕňom v oku.	Fra

ani trochu	
ani máčny mak	Fra

Intonácia *Int*

Peter prišiel	
Peter *prišiel*	Int
Peter prišiel	Int

Básnické figúry

SAPFO by nebola hodná svojho mena, keby sme sa nepokúsili naučiť ju tvoriť aj niektoré básnické a rečnícke figúry. Dobre si všimnite (najmä pri perifráze), v čom spočíva podstata problému pri tvorení figúr: *Počítač sa podobá človeku, ktorý má veľa vedomostí, ale nie je múdry*. Pravidlá tvorenia figúr sa síce naučí, ale aplikuje ich na nesprávnom mieste v nesprávnom čase. Tým demonštruje, že iba *opakuje* po druhých, ale *duch* veci mu uniká.

Litotes *Lit* [+rečn]

zlý výkon	
nie najlepší výkon	Lit
nie veľmi dobrý výkon	Lit

škaredý	
nepekný	Lit

Litotes funguje so sémantickou triedou kvalitatívno-hodnotiacich adjektív. Vzniká vlastne zložením antonymizácie, gradácie a negácie: *Lit* = *NegGraAnt*.

Komparácia *Kom* [+mag,+poet,expr]

čistý	
čistý ako ľalia	Kom

Lexikalizované a konštantné prirovnania (*vraný koník*) sa čerpajú z lexikálno-sémantického slovníka. Pripájajú sa spojkami *ako, sťa, ani*. Intenzitné prirovnanie pripája komponent +mag; spojka môže pripojiť štylistický príznak (*rovný ani svieca, +expr.*; *driek sťa vŕbový prútik, +bás.*).

Program nedokáže rozpoznať prístavkové prirovnanie a pokladá ho za ozajstnú predikáciu: *Predstavil si ju, kráľovnú silvestrovského plesu, v bielych šatách. (Jaroš)*

Perifráza *Pfr*

Bratislava	
mesto na Dunaji	Pfr

Trenčín	
hrad Matúša Čáka	Pfr

V súčinnosti s databankou program dokáže pracovať aj s perifrázami určitého druhu. Robí to tak, že perifrázované meno nahradí hyperonymom a potom k nemu pripojí nejakú informáciu z databanky, ktorá môže stáť v roli prívlastku (obyčajne LOKATÍV, POSESÍV). Pri analýze sa darí úspešne identifikovať denotát perifrázy (*hráči v modrých dresoch → Slovan*). Pri syntéze program generuje aj perifrázy, ktoré postrádajú práve podstatu svojho zmyslu, pretože sú nešpecifické (*Slovan → hráči v dresoch, Bratislava → mesto s radnicou*).

Oxymoron *Oxy*

mŕtvy
živá mŕtvola Oxy

bohatý
chudobný boháč Oxy

Oxymoron sa konštruuje s využitím antoným a deadjektívnej derivácie. Ťažko povedať, či všetky takto mechanicky vzniknuté oxymorony dávajú zmysel (*lacná drahota, prázdna plnosť, zlý dobrák*).

Perisológia *Psg*

zle
zle-nedobre Psg

Perisológia využíva antonymizáciu a negáciu. Je dobre obmedziť ju iba na kvalitatívno-hodnotiace adjektíva, inak vznikajú pomerne nezaujímavé perisológie (*prázdny-neplný, nízky-nevysoký, drahý-nelacný*).

Tautológia *Tau*

div
div divúci Tau

celý
celý celučičký Tau

škoda
škoda preškoda Tau

SAPFO tvorí tautológie tak, že zo slovotvorného hniezda príslušného slova vyberie ešte jeden, dubletný, deminutívny, elatívny alebo iný tvar. Vznikajú zaujímavé aj menej zaujímavé tautológie (*slnce slniečko, starý staručký, iskierka iskrivá, drahý predrahý, Fero Ferinko, divný podivný, strašný strašlivý, drevo drevené).*

Epanalepsia *Epa*

neučí sa
učiť sa neučí Epa

Napísal som, len čo ďalej?
Napísať som napísal, len čo ďalej? Epa

Vzniká anteponovaním prísudkového slovesa v infinitíve. Program nedokáže rozoznať, kedy je epanalepsia vhodná (*Ísť na trh pôjdeš a doniesť donesieš kilo jabĺk.)*

Symploké *Spl*

Dosť ich bolo.
Dosť ich bolo, dosť. Spl

Dá sa tvoriť zopakovaním prvého aktantu na konci vety za čiarkou *(Kam si sa mi podela, kam? Do doliny zaletela, do doliny.)*

Chiazmus *Chi*

Ľúbiš ma ešte?
Ľúbiš ma ešte, ešte ma máš rád? ChiSyn

Chiazmus vzniká zopakovaním aktantov v zrkadlovom poradí. Mechanicky utvorený chiazmus spravidla nemá veľa do seba (*Do mesta ideš, ideš do mesta? Idem kúpiť chlieb, chlieb idem kúpiť.)*

Atrakcia *Atr*

hory a doliny
hory-doly Atr

z očí v oči
zoči-voči Atr

Pri atrakcii ide o zvukovú asimiláciu, dozvuk, spodobenie. Ide o malý počet ustálených spojení vymenovaných v slovníku (*zôkol- vôkol).*

14.

INTELIGENTNÝ NÁVRAT INFORMÁCIE Z DATABANKY

14.1 Problém sémantickej relevancie

Sformulujme problém *inteligentného návratu informácie z databanky v prirodzenom jazyku* nasledovne: Majme veľkú banku dát **D**, obsahujúcu informácie všeobecného charakteru a uložené vo forme viet prirodzeného jazyka (knižnica, archív a pod.). Ďalej máme danú otázku Q, tiež v prirodzenom jazyku. Našou úlohou je algoritmicky porovnať otázku Q s bankou dát **D** a vyznačiť v **D** všetky odstavce, ktoré sú s najväčšou pravdepodobnosťou významovo relevantné otázke Q.

Uvedomme si, že nemáme k dispozícii žiaden *formálny* jazyk, do ktorého by sme mohli databázu **D** a otázku Q preložiť a potom ľahko porovnať. Takýto formálny jazyk je k dispozícii iba v prípade špeciálnych, uniformne štruktúrovaných databáz s dobre definovanou konečnou sémantikou. Ale *jediný* symbolický nástroj, ktorý je schopný pracovať s informáciou ako takou (bez kladenia ohraničení na diskurz), je len prirodzený jazyk.

Porovnávať Q a **D** ako reťazce znakov zrejme nepovedie k cieľu. Problém návratu informácie v prirodzenom jazyku je v tom, že význam, na ktorý sa pýtame, môže byť v **D** skryte prítomný, ale v úplne inej forme ako v Q, alebo môže z **D** vyplynúť až po nejakom druhu logického dôvodenia. V **D** sa môžbyť nachádza podmnožina viet **Q'**, ktorá predstavuje prijateľnú odpoveď na otázku Q, ale my to nedokážeme rozpoznať, lebo v **Q'** je použitá úplne iná lexika a odlišné syntaktické konštrukcie; navyše hľadaná informácia bude rozptýlená na viacerých miestach v **D** a premiešaná s množstvom irelevantnej informácie.

14.2 SAPFO ako informačný systém (Information Retrieval System)

V medziach formálneho systému, ktorý sme opísali vyššie, je SAPFO schopné fungovať ako systém pre návrat informácie z databanky, resp. systém pre zodpovedanie otázok.

Ak je daná dvojica fráz L_1, L_2, systém ich najprv zobrazí na ich faktické reprezentácie $R(L_1)$, $R(L_2)$. Potom sa snaží eliminovať stále viac sémantických diferencií medzi $R(L_1)$, $R(L_2)$, až kým nenájde ich spoločný prienik. Nakoniec

buď	$R(L_1) \cap R(L_2) = \emptyset$,	L_1, L_2 asi nebudú relevantné
alebo	$R(L_1) \cap R(L_2) \neq \emptyset$,	L_1, L_2 môžu byť relevantné

Ak počítač nájde postupnosti formálnych manipulácií

$$T_{11}, T_{12}, \ldots, T_{1n} \text{ a } T_{21}, T_{22}, \ldots, T_{2m}$$

také, že

$$T_{11}T_{12}\ldots T_{1n}L_1 = L_1'$$
$$T_{21}T_{22}\ldots T_{2m}L_2 = L_2'$$

a L_1', L_2' majú neprázdny spoločný prienik, t. j. sú unifikovateľné algoritmom jednoduchého porovnávania so vzorom (simple pattern matching)

$$L_1' \cap L_2' \neq \emptyset,$$

potom L_1 a L_2 *môžu* byť sémanticky relevantné. Postupnosť

$$T_{2m}^{-1}\ldots T_{21}^{-1}T_{11}\ldots T_{1n}$$

je syntakticko-sémantický most medzi L_1 a L_2 a sémantická diferencia L_1- L_2 je daná zoznamom komponentov asociovaných s použitými manipuláciami. Algoritmus je založený na hypotéze, že ak je počet syntaktických a sémantických diferencií medzi L_1, L_2 malý, L_1 a L_2 si budú s veľkou pravdepodobnosťou aj významovo blízke a naopak: ak sú dve frázy blízke významovo, mal by byť počet syntakticko-sémantických diferencií medzi nimi ohraničený. Sémantická diferencia predstavuje v sémantickom poli to isté, čo *Hammingova vzdialenosť* v teórii kódovania.

Takéto prehľadávanie je, pravda, do istej miery 'slepé'. Videli, sme, že systém je sotva kompetentný rozhodnúť, ktoré manipulácie a v akom poradí je prípustné aplikovať. Vznikajú aj parafrázy, ktoré sú významovo nezmyselné alebo nemajú žiadne denotáty v reálnom svete. Avšak tieto parafrázy sú vždy nezmyselné akosi *pozdĺž logiky prirodzeného jazyka* a podobajú sa na výpovede detí vo veku okolo piatich rokov, keď si osvojujú pravidlá jazykového prejavu (*strelič, veľrybička, rybolovenie).* Prehľadávanie sa teda deje pozdĺž logiky prirodzeného jazyka, nie pozdĺž iných formálnych štruktúr (ako napríklad predikátový počet prvého rádu), ktoré sú prirodzenému jazyku úplne cudzie. Ak by sa informačnému systému podarilo vypísať z databanky na obrazovku vždy povedzme 100 viet, z ktorých 10 by bolo skutočne relevantných výzve používateľa, a 90 by bolo balastom, mohli by sme systém pokladať za úspešný. Používateľ ľahko prezrie stovku viet a vyberie si pre seba užitočnú informáciu, ale celú knižnicu by neprezrel ani do konca života. Teda ak sa hľadaná informácia nenachádza práve tam, kde by podľa jednoduchého knižničného katalógového kľúča mala patriť, používateľ sa k nej nikdy nedostane, napriek tomu, že je v knižnici prítomná.

Predbežný experiment so SAPFO ako informačným systémom naznačuje, že 80 % relevantných textov systém nerozpoznáva, pretože premostenie k nim vyžaduje využívať poznatky o reálnom svete, nie len poznatky jazykové. Náš formálny aparát je teda nedostatočný na to, aby mohol byť úspešný vo funkcii informačného systému.

Jazyková algebra by však mohla úspešne dopĺňať a tvoriť účelnú kombináciu s metódami štatistickými a metódami logickej inferencie. Experimenty s aplikovanými informačnými systémami na štatistickej báze ukazujú, že takýto prvok by mohol byť žiadúci (Salton, 1968, 1981).

14.3 Výpočet sémantickej relevancie. Príklad

Predpokladajme, že máme otázku:

Q = *Nakupoval Jano cenné veci?*

a databáza **D** obsahuje (okrem iného) aj vetu:

A = *Pavol vedel o Petrovom predaji diamantu Janovi.*

Po transmanipulácii Q a A dostaneme:

Pavol vedel o Petrovom predaji diamantu Janovi.	$/Nol^{-1}Nor^{-2}$	
Pavol vedel, že Peter predal diamant Janovi.	$/Kon^{-1}$	
Pavol vedel, že Jano kúpil diamant od Petra.	/Abs	
Pavol vedel, že Jano kúpil drahokam od Petra.	$/Kpz^{-1}$	
*Pavol vedel, že **Jano kúpil drahý kameň** od Petra.*		= A'

Nakupoval Jano cenné veci?	$/Que^{-1}$	
Jano nakupoval cenné veci.	$/Reo^{-1}$	
Jano kupoval cenné veci.	/Asp	
Jano kúpil cenné veci.	/Der	
Jano kúpil cennosti.	$/Plu^{-1}$	
Jano kúpil cennosť.	/Syn	
Jano kúpil drahokam.	$/Kpz^{-1}$	
Jano kúpil drahý kameň.		= Q'

Máme:

$Kpz^{-1}SynPlu^{-1}DerAspReo^{-1}Que^{-1}Q = Q'$
$Kpz^{-1}AbsKonNor^{-2}Nol^{-1}A = A'$
$Q' \leq A'$
$Q \leq QueReoAsp^{-1}Der^{-1}PluSyn^{-1}KpzKpz^{-1}AbsKonNor^{-2}Nol^{-1}A$

Postupnosť $ReoAsp^{-1}Der^{-1}PluSyn^{-1}AbsKonNor^{-2}Nol^{-1}$ je syntakticko-sémantický most medzi Q a A. Členy postupnosti svedčia o druhu a rozsahu rozdielov medzi

otázkou Q a odpoveďou A. *Nor*^{2}, *Nol*, *Der* spôsobili iba formálne zmeny. *Kon* zmenila vetnú perspektívu. *Reo*$^{-1}$ eliminovala orientáciu deja *kupovať/nakupovať* (+ex), *Asp* zmenila vid z nedokonavého na dokonavý (+perf), *Plu*$^{-1}$ pozmenila vetné členy v čísle (-plu). *Abs* zovšeobecnila význam A, pričom A zostala iba špeciálnym prípadom A' (A ≤ A'). *Syn* spôsobila bližšie neurčené zmeny v medziach lexikálnej synonymie (Q ≈ Q'). Výsledná sémantická rovnica je:

$$Q \approx A \text{ +ex -perf -plu}$$

V tomto prípade, ako vidíme, je Q úplne relevantnou odpoveďou na A aj po aplikácii 12-tich formálnych manipulácií.

15.

FREKVENČNÁ ANALÝZA SELEKCIA PARAFRÁZ PODĽA PRAVDEPODOBNOSTI

15.1 Kombinatorická explózia počtu parafráz

Pri generovaní sesterských povrchových štruktúr z jednej hĺbkovej štruktúry existuje značná rozmanitosť a možnosť voľby. V odseku 12.22 sme ukázali, že horný odhad počtu transmutovaných parafráz *i*-teho rádu v našom systéme je

$$P_{PAR} \leq 2.66^{i}.n^{i+2}.n!.3^{4nl}$$

kde *n* je počet vrcholov a *l* počet hrán v sieti. Číslo *66* je počet komponentov významu a číslo *3* znamená, že faktické pády môžu byť sémanticky najviac trojperspektívne.

Predstavme si napríklad päťslovnú vetu, v ktorej každý vzťah by sa dal vyjadriť troma formami, dva vzťahy by boli dvojperspektívne a každý vrchol by sa dal aspoň jedným spôsobom mutovať. Takáto veta by dávala 5 intonačných parafráz, okolo 10 transfokácií, 81 transformácií, 4 transfigurácie a 5 transmutácií prvého, 25 druhého, 125 tretieho rádu. Spolu 212.463.000 parafráz.

15.2 Problém voľby

Človek ani stroj s takýmto množstvom parafráz nedokáže pracovať. Okrem toho nie všetky parafrázy sú rovnako dobre prijateľné; je nevyhnutné rozlišovať medzi nimi z hľadiska kvality. Navyše rozličné parafrázy sú vhodné v rozličných kontextoch.

V hovorenom a písanom slove sa výber vhodných vyjadrovacích prostriedkov uskutočňuje na základe celého radu komplexných faktorov: kontext, konsituácia, komunikatívny zámer a citový postoj hovoriaceho k obsahu výpovede, predpokladaný stupeň informovanosti adresáta, celková štylistická forma slovesného prejavu, spoločenské postavenie a vzťah medzi hovoriacim a počúvajúcim, dostatok či nedostatok času a mnoho ďalších faktorov.

Pretože väčšina z týchto vonkajších faktorov nie je súčasťou nášho modelu, obmedzíme sa iba na určitý jednodimenzionálny spôsob ohodnotenia parafráz: na množine parafráz danej frázy zavedieme usporiadanie pomocou číselných koeficientov tak, aby málo pravdepodobné, vágne parafrázy so zložitou konštrukciou, nezvyčajnou lexikou alebo málo pravdepodobným významom dostali číselne nízke ohodnotenie. Naopak typické, zvučne znejúce parafrázy, aké sa v slovesných preja-

voch s najväčšou pravdepodobnosťou vyskytnú, musia obdržať vysoký hodnotiaci koeficient.

Týmto spôsobom sa pokúsime oddeliť najkvalitnejšie parafrázy od veľkého množstva balastu.

15.3 Fuzzy-logika. Spracovanie neistoty

Aplikovaná matematika je v praxi často konfrontovaná s úlohami, ktoré sa dajú len ťažko zachytiť a popísať pomocou jednoznačných, určitých pojmov.

Jedným z matematických prístupov k spracovaniu neistoty je tzv. neostrá (fuzzy) logika a neostré množiny (fuzzy sets; Fuzzy, 1983). Jednotlivé prvky môžu do fuzzy-množiny patriť alebo nepatriť s určitou pravdepodobnosťou; sylogizmy vo fuzzy-logike môžu byť s určitou pravdepodobosťou *p* správne a zároveň nesprávne s pravdepodobnosťou *1-p*.

Spracovanie prirodzeného jazyka tiež patrí k tomuto typu úloh. Jednotlivé gramatické, syntaktické, slovotvorné a štylistické javy fungujú len s určitou pravdepodobnosťou, a z každého pravidla existujú výnimky. Jednotlivé prvky jazyka možno zaradiť do určitých tried a paradigiem, ale vždy iba s určitou pravdepodobnosťou (a s určitou pravdepodobnosťou do nich patriť nemusia).

V tejto kapitole navrhneme spôsob, ktorým možno „hmlistý" prístup uplatniť pri počítačovom spracovaní jazyka.(O chápaní jazyka ako pravdepodobnostného systému porov. aj Nalimov, 1979).

15.4 Pravdepodobnostné váhové koeficienty a ich štatistické určovanie

Budeme postupovať takto: Ku každej elementárnej slovníkovej a morfosyntaktickej operácii, ktorá sa pri externalizácii hĺbkovej štruktúry uplatňuje, priradíme *váhový koeficient* vyjadrujúci pravdepodobnosť, s akou sa daný vyjadrovací prostriedok používa oproti alternatívnym vyjadrovacím prostriedkom, ktoré mu konkurujú. Takéto koeficienty získame štatisticky na reprezentatívnom korpuse slovenských textov.

Napríklad v štylistickom slovníku pribudnú ku každej položke číselné ohodnotenia:

štyl(väzenie,neutr,87%).
štyl(trestnica,kniž,2%).
štyl(žalár,kniž,3%).
štyl(temnica,básn,2%).
štyl(árešt,arch,1%).
štyl(basa,hovor,5%).
štyl(loď,neutr,90%).
štyl(koráb,básn,10%).

Suma ohodnotení vzájomne si konkurujúcich štylistických pendantov musí byť vždy spolu rovná 100 %. Koeficienty sú priamo úmerné percentuálnemu zastúpeniu daných lexikálnych jednotiek v korpuse textov.

Analogicky zorganizujeme frazeologický aj synonymický slovník. Koeficienty vyjadrujú pravdepodobnosť, s akou frazeologické a synonymické jednotky nahrádzajú základné vyjadrenie:

frazeol(živor-, [tr- PAC» bied-] ,4%). *(trieť biedu)*

frazeol(živor-, [preklad- PAC» psot- ORI» plec- DES» plec-] ,1%). *(psotu z pleca na plece prekladať)*

symbol(výnimk-, [vran- ATR» biel-] ,3%). *(biela vrana)*

synon(plak-, [ron- PAC» slz- $^{+pl}$] ,4%). *(roniť slzy)*

synon(sopečn-,vulkanick-,50%).
synon(plak-,bedák-,5%).
synon(plak-,nariek-,7%).
synon(plak-,horekov-,2%).

Ak tento systém použijeme aj v lexikálno-sémantickom slovníku, pomôže nám to podchytiť problém polysémie. Váhové koeficienty popri dištinktívnych sémantických príznakoch významne prispejú aj k odlíšeniu jednotlivých lexií:

subst(srdce$_1$,korp,44%). *(telesný orgán)*
subst(srdce$_2$,gnoz,44%). *(ľudská schopnosť cítiť)*
subst(srdce$_3$,lok,8%). *(srdce krajiny)*
subst(srdce$_4$,kon,4%). *(srdce zvonu)*

subst(loď$_1$,ins,97%). *(plavidlo)*
subst(loď$_2$,lok,3%). *(loď kostola)*

subst(Napoleon,osoba,100%)

Váhové koeficienty možno priradiť aj gramatickým operáciám. Pretože skloňovanie a časovanie sú spravidla jednoznačné, väčšina gramatických tvarov dostane váhový koeficient 100 %. Iba kodifikované dvojtvary *(čerešňí, čerešieň)* alebo slová migrujúce od jedného vzoru k druhému *(od lode, od lodi)* sa o svoje percentuálne zastúpenie podelia:

skloňovanie(čerešňa,ruža,bez_alternácie,60%).
skloňovanie(čerešňa,višňa,vlož_ie,40%).
skloňovanie(loď,kosť,_,50%).
skloňovanie(loď,dlaň,_,50%).

Nakoniec priradíme váhové koeficienty všetkým syntaktickým konštrukciám a sémantickým kolokáciám. V súlade s našou filozofiou o neoddeliteľnosti syntaxe a sémantiky to urobíme tak, že váhový koeficient priradíme každému kolokačnému pravidlu, ktoré spája formu s významom. Koeficient bude vyjadrovať pravdepodobnosť, s akou sa daný význam vyjadruje danou formou - v konkurencii s ostatnými formami, ktorými možno tento význam tiež vyjadriť. V praxi to znamená priradiť jednotlivým riadkom všetkých valenčných funkcií (VAL, LVAL, RVAL, CVAL, DVAL, AVAL) ich pravdepodobnostnú početnosť. Napríklad riadky

lval((adj,indprivl,ov),(sub,osoba),(sub,kon,_),_,ori,25%).
lval((adj,akvzt,ský),(sub,_),(sub,kon,_,_),_,ori,25%).
rval((sub,kon),_,0,gen,(sub,osoba),_,ori,25%).
rval((sub,kon),_,od,gen,(sub,osoba),_,ori,25%).

vyjadrujú, že vzťah pôvodu (originatív) v hĺbkovej štruktúre

[posol-
ORI» kráľ-]

možno vyjadriť štvorakým spôsobom a to približne s rovnakou pravdepodobnosťou:

kráľov posol
káľovský posol
posol kráľa
posol od kráľa

Všetky riadky slovesných intenčných rámcov dostanú svoju pravdepodobnostnú váhu:

(01) val(manuálny_proces_22,0,nom,živ,ags,40%).
(02) val(manuálny_proces_22,0,ins,živ,ags,10%).
(03) val(manuálny_proces_22,0,nom,ins,ins,10%).
(04) val(manuálny_proces_22,0,ak,_,pac,50%).
(05) val(manuálny_proces_22,0,ins,ins,ins,40%).
(06) val(manuálny_proces_22,na,lok,ins,ins,10%).
(07) val(manuálny_proces_22,pomocou,gen,ins,ins,5%).
(08) val(manuálny_proces_22,za_pomoci,gen,ins,ins,5%).

(09) lval((adj,indprivl,ov),(sub,osoba),(sub,dej,_),(sl,I),SP,X)
:- val(I,_,_,živ,SP),((SP=ags,X=20%);(SP=pac,X=10%)).

(10) rval((sub,dej),(sl,I),0,gen,(sub,osoba),_,SP,20%)
:- val(I,_,_,živ,SP).

(11) cval(lk(sub,ins,_),rk(sl,I,_),kk(sub,osoba,ec),[ags,3% / pac,2%]).

(12) cval(lk(sl,I,_),rk(sub,Se,_),kk(sub,osoba,''), [ags,0.3% / pac,0.2%]).

(13) cval(lk(sub,lok,_),rk(sub,os,ec),kk(sub,os,_),[pac,4% / ins,3%]).

(14) dval((sub,osoba,_,ec),(sl,I),ags,6.7%).

(15) dval((sub,osoba,_,anec),(sl,I),pac,3.8%).

(16) dval((sub,ins,_,dlo),(sl,I),ins,7%).

Všimnite si, že riadky 1, 2, 9, 10, 11, 12 a 14, prostredníctvom ktorých sa vyjadruje AGENS manuálneho procesu, dávajú spolu 100 %. AGENS býva spravidla vyjadrený ako vetný podmet v nominatíve (*otec stavia*), zriedkavejšie v pasívnom inštrumentáli (*postavený Rimanmi*). S pravdepodobnosťou $^1/_5$ sa agens realizuje ako osobné privlastňovacie prídavné meno (*otcovo pílenie*) alebo ako genitívny prívlastok (*pílenie otca*). Ďalšia možnosť je spraviť z agensa deverbatívny derivát (*lovec, pracovník*).

S minimálnou pravdepodobnosťou prichádza do úvahy realizácia AGENSa vo forme kompozita (3.3 %). Riadky 11 a 12 vyjadrujú aj to, že ak sa agens má vyjadriť formou zloženého slova, bude to s desaťnásobne väčšou pravdepodobnosťou (3 %) podľa kompozitného vzorca 'substantívum-sloveso' (*drevorubač, lodivod*) než podľa vzorca 'sloveso-substantívum' (*tlčhuba, držgroš, valibuk*), ktorý sa uplatňuje len s pravdepodobnosťou 0.3 %.

Na týchto istých kompozitách sa zároveň podieľa PACIENS svojimi 2.2 percentami, ktoré spolu s riadkami 4, 9, 10, 13 a 15 dávajú plných 100 %. Podiel PACIENSa v kompozite typu *lodivod* vychádza odlišný ako pri AGENSe, pretože celkové pravdepodobnostné rozloženie výskytu PACIENSa je odlišné. Môže sa napríklad vyskytovať v spoločných kompozitách aj s INŠTRUMENTom (*kávovar*).

INŠTRUMENT sa vyjadruje prostredníctvom valenčných riadkov 3, 5, 6, 7, 8, 10, 13 a 16, ktoré sú opäť volené tak, aby ich súčet bol 100.

Riadok 9 zachycuje, že v pozícii ľavostranného prívlastku sa AGENS vyskytuje dvakrát častejšie ako PACIENS (*zločincov útek* oproti *zločincovo zatknutie*).

Riadky 15 a 16 zachycujú, že forma derivátu sa často používa na vyjadrenie nástroja (*varič, sekačka*), ale iba zriedka na vyjadrenie objektu (*trestanec, skamenelina*).

15.5 Usporiadanie parafráz podľa pravdepodobnosti

Váhové ohodnotenie jednotlivých elementárnych slovníkových, morfosyntaktických a sémanticko-väzobných operácií nám umožní zaviesť na množine parafráz

lineárne usporiadanie. Celková váha parafrázy bude daná normovaným súčinom elementárnych váh všetkých operácií použitých pri jej generovaní.

Príklad. Parafrázujme vetu

Jano ide loďou do Indie.

Nech slovník a valenčné funkcie sú ohodnotené takto:

lex(loď$_1$,ins,97).
lex(loď$_2$,lok,3).
lex(ísť$_1$,translačný_proces_1,90).
lex(ísť$_2$,limitné,10).

synon(ísť$_1$,cestovať,30)
synon(loď$_1$,koráb,10)

val(translačný_proces_1,0,nom,živ,ags,50).
val(translačný_proces_1,0,ins,ins,ins,40)
val(translačný_proces_1,na,lok,ins,ins,20)
val(translačný_proces_1,pomocou,gen,ins,ins,5)
val(translačný_proces_1,za_pomoci,gen,ins,ins,4)
val(translačný_proces_1,prostredníctvom,gen,ins,ins,3)
val(translačný_proces_1,0,ins,lok,via1,20).
val(translačný_proces_1,cez,ak,lok,via1,40).
val(translačný_proces_1,naprieč,inš,lok,via1,5).
val(translačný_proces_1,do,gen,lok,des1,60).
val(translačný_proces_1,doprostred,gen,lok,des1,5).
val(translačný_proces_1,naprostred,gen,lok,des1,5).

rval((sub,mov,_),_,do,gen,(sub,lok,_),_,des1,10).
rval((sub,nemov,_),_,do,gen,(sub,lok,_),_,des1,5).

Analyzátor skonštruuje štyri hĺbkové štruktúry, ktoré zodpovedajú štyrom rozdielnym významom (čítaniam) vety:

1. Jano sa plaví loďou do Indie. (čím)
2. Jano ide loďou (kostola) do Indie. (kadiaľ)
3. Jano ide na lodi, ktorá ide do Indie.
4. Jano ide loďou (kostola), ktorá je (nasmerovaná) do Indie.

Ľudský subjekt by v kontexte zvyšné tri významy (okrem prvého) podvedome prehliadol. Počítaču nemožno zazlievať, že uvažuje všetky štyri možnosti. Dostal ich iba dôslednou aplikáciou sémantických pravidiel, ktoré sme do gramatiky systé-

mu vložili. Všetky štyri alternatívy sú logicky možné, nie sú však rovnako pravdepodobné.

Generátor teda bude tvoriť parafrázy od všetkých štyroch hĺbkových štruktúr - aplikáciou rozmanitých vyjadrovacích prostriedkov. Nakoniec parafrázy zoradí podľa ich váhových koeficientov. Dostaneme takýto obraz:

Jano ide loďou do Indie.	
Jano ide do Indie loďou.	0.1080
Jano ide do Indie na lodi.	0.0540
Jano ide loďou, ktorá ide do Indie.	0.0450
Jano cestuje loďou do Indie.	0.0324
Janova cesta loďou do Indie.	0.0216
Jano ide do Indie pomocou lode.	0.0135
Jano ide cez loď (kostola) do Indie.	0.0120
Jano ide do Indie pomocou lode.	0.0108
Jano ide korábom do Indie.	0.0108
Jano ide loďou doprostred Indie.	0.0090
Jano ide loďou naprostred Indie.	0.0090
Jano ide do Indie prostredníctvom lode.	0.0081
Jano ide loďou (kostola) do Indie.	0.0060
Jano ide pomocou lode, ktorá ide do Indie.	0.0045
Jano cestuje korábom do Indie.	0.0032
Jano ide naprieč loďou (kostola) do Indie.	0.0015
Jano ide cez loď (kostola nasmerovanú) do Indie.	0.0010
Jano ide pomocou lode doprostred Indie.	0.0010
Jano ide loďou (kostola nasmerovanou) do Indie.	0.0005

Pod čiarou sa začínajú tie parafrázy, ktorých váha poklesla pod jednu desatinu váhy prvej (t. j. najpravdepodobnejšej) z nich.

Vidíme, že do čela sa dostali celkom „rozumné" parafrázy. Zatiaľ čo parafrázy obsahujúce viacero prvkov s malou pravdepodobnosťou (pod 10 %), prepadnú pod čiaru medzi tisícky parafráz s pravdepodobnosťou blízkou nule.

Týmto spôsobom sa darí vyradiť maximálne množstvo druhoradých a treťoradých parafráz, ktoré vznikajú kombinačnými možnosťami vyjadrovacích prostriedkov.

Táto selekcia je nemenej dôležitá aj z hľadiska výpočtovej zložitosti. Ak totiž váhový koeficient nejakej parafrázy už počas výpočtu poklesne pod $^1/_{10}$ maximálnej priebežne zistenej váhy, môžeme tvorbu takej parafrázy okamžite prerušiť. V praxi to pre 90 % parafráz znamená, že nebudú vôbec vygenerované.

Pravda, tento mechanizmus je primálo citlivý ku skutočnému významu viet, než aby dokázal pracovať bezúhonne. Napríklad vo vete

Filmový producent si vzal hviezdu.

sú viac pravdepodobné práve nižšie ohodnotené lexie slov *vzať si* (premiestniť, oženiť sa) a *hviezda* (nebeské teleso, slávna osoba), a to vďaka selekčnému efektu, ktorý vzniká interakciou konotačných polí slova *film* a *hviezda*.

16.

VÝPOČTOVÁ ZLOŽITOSŤ ALGORITMOV

16.1 Zložitostné triedy

Z hľadiska praktickej aplikovateľnosti je okrem otázky, či daný algoritmus vedie k správnemu výsledku, nemenej dôležitá aj otázka jeho výpočtovej zložitosti; t. j. aké veľké množstvo času, pamäťového priestoru a energie algoritmus na výpočet potrebuje? Existujú totiž algoritmy, ktoré vedú k správnemu výsledku, ale v nekonečnom čase, alebo algoritmy, ktoré sú síce konečné, ale množstvo potrebných operácií je také veľké, že na výsledok by sme museli čakať roky a tisícročia.

V teórii zložitosti sa časová a priestorová zložitosť vyjadrujú ako funkcie dĺžky vstupu *n:*

$$T(n) = F_1(n); \qquad S(n) = F_2(n)$$

F je spravidla niektorá z elementárnych funkcií (alebo je ich lineárnou kombináciou), ktorá asymptoticky vyjadruje približný rast počtu potrebných operácií (či pamäťových buniek) s rastúcou veľkosťou úlohy. Pre veľké *n* je potom rozhodujúci iba typ dominantného člena funkcie, ktorý všetky ostatné členy prerastie.

Základná zložitostná škála vyzerá takto:

$$\log n \le n \le n.\log n \le n^k \le n^{\log n} \le k^n \le n! \le n^n$$

Za algoritmy vypočítateľné v reálnom čase sa pokladajú algoritmy so zložitosťou polynomiálnou (n^k) alebo nižšou. Existuje dôležitá trieda úloh, ktoré sú riešiteľné v exponenciálnom čase k^n, doteraz sa však pre ne nepodarilo nájsť algoritmus s menšou zložitosťou a nie je jasné, či takýto algoritmus vôbec existuje. Označuje sa ako zložitostná trieda *NP* (non-polynomial) a každá úloha, ktorá do tejto triedy patrí, alebo je na niektorú z týchto úloh transformovateľná, sa nazýva *NP-ťažká.*

Časová zložitosť	*Veľkosť úlohy*		
	10	50	100
n^3	0.001 sek.	0.125 sek.	1.0 sek.
2^n	0.001 sek.	35.7 rokov	10^{15} storočí

Ilustratívne porovnanie polynomiálneho a exponenciálneho rastu zložitosti. Na počítači, ktorý vykonáva 1 milión (10^6) operácií za sekundu, vyrieši kubický (n^3) aj exponenciálny (2^n) algoritmus úlohu o veľkosti 10 za rovnaký čas (jednu tisícinu sekundy). No zatiaľ čo kubický algoritmus vyrieši ešte aj úlohu o veľkosti 100 v reálnom čase (za 1 sekundu), exponenciálny algoritmus by na ňu potreboval miliardy rokov.

16.2 Zložitosť vyhľadávania v slovníku

Slovník SAPFO je organizovaný do tvaru vyváženého stromu (balanced tree). Ako je známe (Hopcroft - Ullman, 1979), zložitosť k-árneho prehľadávania usporiadaného zoznamu je logaritmická:

$$T(n) = \log_k n,$$

kde *k* je počet vetvení b-stromu a *n* je veľkosť zoznamu. S každým zdvojnásobením objemu slovníka sa čas vyhľadávania zvýši iba lineárne.

Rast slovníka teda neohrozuje systém z hľadiska spotreby výpočtového času.

16.3 Zložitosť morfologickej analýzy

V kapitole 3 sme rozobrali kombinatorické a zložitostné problémy tvaroslovnej analýzy. Ukázali sme, že pomocou konečného automatu sa nám zložitosť podarilo zredukovať na minimálnu možnú mieru, ktorá sa rovná vlastnej zložitosti slovenskej morfológie a nie je už ďalej redukovateľná. Konečný automat pracuje v lineárnom čase (Hopcroft - Ullman, 1979), pričom časová náročnosť sa rovná dĺžke vstupu:

$$T(n) = n$$

Náš automat pracuje nedeterministicky lineárne, t. j. v najhoršom prípade v polynomiálnom čase:

$$T(n) = n^k$$

Počet vyhľadávaní v slovníku, ktoré treba vykonať

$$P = P_{suf} * P_{pref} * P_{alt}$$

je daný súčinom počtu prípon, ktoré možno oddeliť od konca slova, počtu predpôn oddeliteľných od začiatku slova a počtu spôsobov, koľkými je možné koreň alternovať.

Zložitosť nášho algoritmu sa teda rovná vlastnej zložitosti problému ako takého a nedá sa ďalej zefektívniť, ak to nemá byť za cenu neadekvátneho nárastu pamäťového priestoru alebo energie.

Hoci väčšina rýchlych morfologických analyzátorov je realizovaná ako konečné automaty, G. E. Barton a kol. (1987) na príklade dvojúrovňového morfologického modelu L. Karttunena (1983) ukázali, že problém morfologickej analýzy je vo všeobecnosti exponenciálne zložitý. Obídenie časovej exponenciality sa teda dosahuje len vďaka tomu, že časová exponencialita sa transformuje do exponenciálnej náročnosti na pamäťový priestor. Počet stavov výsledného konečného automatu je potom úmerný kombináciám všetkých možností, ktoré sa pri tvaroslovnej analýze môžu vyskytnúť.

V jazykoch s chudobnejšou morfológiou, kde počet morfologických javov je ohraničený na desiatky alebo stovky, možno uvedený postup aplikovať bez problémov. Vidíme, že ešte aj pre slovenčinu je to čiastočne možné: automat SAPFO má cca 400 stavov. No v jazykoch s bohatou aglutináciou (turečtina, fínština), kde počet tvarov od jednej lexémy ide do tisícov, alebo v inflexných jazykoch (arabčina), v ktorých fungujú vokalické zhody na dlhé vzdialenosti, horeuvedený postup nie je možný, a skutočná exponenciálna povaha problému sa prejaví.

Aj v jazykoch s produktívnou kompozíciou môže morfologická analýza predstavovať netriviálny problém. M. Kay a R. Kaplan (1982) ilustrujú takúto situáciu na nemeckom slove

Lebensversicherungsgesellschaftsangestellte

(= zamestnanec spoločnosti pre životné poistky).

16.4 Zložitosť syntaktickej analýzy

E. S. Ristad a R. C. Berwick (1989) ukázali, že každá formálna gramatika, ktorá uvažuje kategoriálne zhody na neohraničené vzdialenosti a lexikálnu viacznačnosť, je NP-ťažká. G. E. Barton Jr. a kol. (1987) ďalej dokazujú, že CG (categorial grammar), LFG (lexical-functional grammar) a GPSG (generalized phrase structure grammar) napospol nie sú riešiteľné v polynomiálnom čase. B. Carpenter (1991) dokonca dokázal, že generatívna sila HPSG (head-driven phrase structure grammar) je ekvivalentná triede všetkých rekurzívne vyčísliteľných jazykov.

Regulárne jazyky sú síce rozpoznateľné v čase O(n) a bezkontextové v čase $O(n^3)$, svojou generatívnou kapacitou však nepokrývajú prirodzený jazyk. Objavil sa rad pokusov o rozšírenie bezkontextového formalizmu alebo reštrikciu kontextových gramatík tak, aby spĺňali požiadavku polynomiálnosti (Rounds, 1988). K. Vijayashankar a A. K. Joshi (1985) navrhli algoritmus $O(n^6)$ pre TAG (Tree Adjoining Grammars).

Pre nás je dôležitý výsledok E. S. Ristada a R. C. Berwicka (1989), podľa ktorého sa dá usudzovať, že gramatika SAPFO je nepolynomiálne zložitá. Ich výsledok

by však zároveň znamenal, že problém analýzy prirodzeného jazyka ako taký je viac než polynomiálne ťažký. Keďže kategoriálne závislosti na dlhé vzdialenosti a lexikálna viacznačnosť sú sotva postrádateľnou súčasťou prirodzeného jazyka, znamenalo by to, že každá iba polynomiálne zložitá gramatika by musela škálu jazykových fenoménov zachytávať len neúplne.

Na druhej strane je tu rad faktorov, ktoré svedčia v prospech toho, že klasická teória zložitosti sa dá na prirodzený jazyk aplikovať iba s výhradami. Tieto dávajú nádej, že úlohy analýzy prirodzených jazykov budú prakticky riešiteľné v reálnom čase, napriek tomu, že patria do triedy exponenciálnych, čiže v reálnom čase neriešiteľných úloh:

a) uvedené odhady platia asymptoticky, t. j. pre veľké *n*. Pri malom *n* sa exponenciálna funkcia nemusí správať horšie, ba môže byť dokonca efektívnejšia od polynomiálnej (napr. $2^n \leq n^{12}$, až po n=70). V prirodzenom jazyku, kde *n* vyjadruje dĺžku vety alebo počet významov či subkategoriálnych príznakov jedného slova, ide práve o takýto prípad.

b) horeuvedené zložitostné charakteristiky sú horným odhadom, t. j. platia „v najhoršom prípade". Priemerná časová zložitosť sa môže od horného odhadu značne líšiť. Gramatika SAPFO, ako každá dostatočne prepracovaná a senzitívna gramatika, dokáže totiž vďaka synergickému efektu (ods. 0.11) a reštrikčným pravidlám skoro vylúčiť veľké množstvo kombinácií, hoci „krajný najhorší prípad" nemožno nikdy vylúčiť, a teda horný odhad musí zostať exponenciálny. Reprezentantom takéhoto krajne kombinatoricky náročného prípadu je napríklad anglická veta *Buffalo buffalo buffalo*, ktorá, hoci je iba trojslovná, predstavuje značnú námahu aj pre ľudské subjekty. Jazyk má pravdepodobne určitú informačnú entropiu (E<<1), ktorá zaručuje, že podobné vety budú vždy v menšine, inak by jazyk bol ťažko zrozumiteľný aj pre ľudí.

16.5 Ako súvisia zložitosť syntaxe, morfológie a slovotvorby?

Všimnime si, že veta *Buffalo buffalo buffalo* je možná iba v angličtine, kde jedno slovo môže byť slovesom, podstatným aj prídavným menom zároveň, a táto morfologická amorfnosť je štatisticky častým javom. Ide teda o akýsi patologický prípad, ktorý je spôsobený atrofiou morfológie a jeho výsledkom je extrémne zaťaženie syntaxe. G. Altmann a R. Köhler (1991) študovali vzájomnú prepojenosť a kvantitatívne vzťahy medzi jednotlivými podsystémami prirodzených jazykov. Ukazuje sa, že zjednodušenie jedného jazykového subsystému si obyčajne vyžaduje zvýšené zaťaženie iného, pričom jazykový systém ako celok sa snaží minimalizovať celkové energetické výdavky. Predpokladajme, že celková zložitosť jazyka sa rovná určitej konštante. Potom si rozloženie zložitosti medzi troma dominantnými jazykovými subsystémami môžeme predstaviť ako spojené nádoby

$$z_{celk} = z_1 + z_2 + z_3 = \text{konšt.},$$

kde z_1 je zložitostný podiel morfológie, z_2 derivatológie a z_3 syntaxe.

Ak predpokladáme, že morfológia, derivatológia a syntax sa dajú riešiť relatívne samostatne (až na tie nejednoznačnosti, ktoré sa prenášajú nerozhodnuté do ďalších subsystémov), potom počet nejednoznačností vo vete bude ležať niekde medzi súčtom a súčinom počtu nejednoznačností jednotlivých subsystémov:

$$k^{z1} + k^{z2} + k^{z3} \leq N_{celk} \leq k^{z1} * k^{z2} * k^{z3} = k^{z1+z2+z3} = \text{konšt.}$$

V ideálnom prípade by sa rovnal iba súčtu. Súčtová funkcia však nadobúda minimum pri hodnotách

$$z_1 = z_2 = z_3 = {}^1/_3 . z_{celk}$$

To znamená, že ideálny jazyk z hľadiska energetickej náročnosti na analýzu by bol jazyk, v ktorom sú všetky tri zložky, t. j. tvaroslovie, slovotvorba a skladba harmonicky rozvinuté a rovnomerne sa podieľajú na celkovej štruktúrovanosti jazyka.

Naše výpočty sú, pravda, zjednodušené. Uvedený výsledok však zodpovedá experimentálnym výsledkom so systémom SAPFO: bohatý tvaroslovný a slovotvorný systém nie sú prekážkou, ale naopak oporou pre syntaktickú analýzu.

16.6 Suboptimálne riešenia

Iný prístup, ktorý sa s úspechom uplatňuje v ťažkých aplikačných úlohách, je hľadanie suboptimálnych riešení. To znamená, že dáme prednosť riešeniu, ktoré je iba blízke optimálnemu, ale zato náklady na jeho výpočet sa značne znížia. Pre niektoré klasické exponenciálne zložité úlohy, ako napr. problém obchodného cestujúceho, takéto algoritmy existujú. M. Garey a D. Johnson (1979) predviedli pre istý NP-úplný problém rýchly heuristický algoritmus, ktorý zaručuje riešenie nie o viac ako 22 % horšie od optimálneho.

Tento prístup sme uplatnili aj v systéme SAPFO zavednením váhových koeficientov ku každej elementárnej gramatickej operácii (15.4). Ušetríme tým množstvo výpočtového času. Cena, ktorú za to musíme zaplatiť, je nebezpečie, že medzi výsledné analýzy a parafrázy sa dostane nejaká nežiadúca, resp. že niektorá žiadúca spomedzi nich vypadne.

16.7 Reálny čas výpočtu

Zaujímavé je porovnať teoretické výsledky so skutočne nameranými číselnými hodnotami.

Výpočtový čas syntaktickej analýzy sme merali na vzorke 100 viet obsahujúcich všetky druhy syntaktických konštrukcií (vrátane takých, ako doplnok, prístavok, prechodníky, infinitívne a participálne konštrukcie, podradené vety, rozvité menné frázy so stupňom rekurzie až 4).

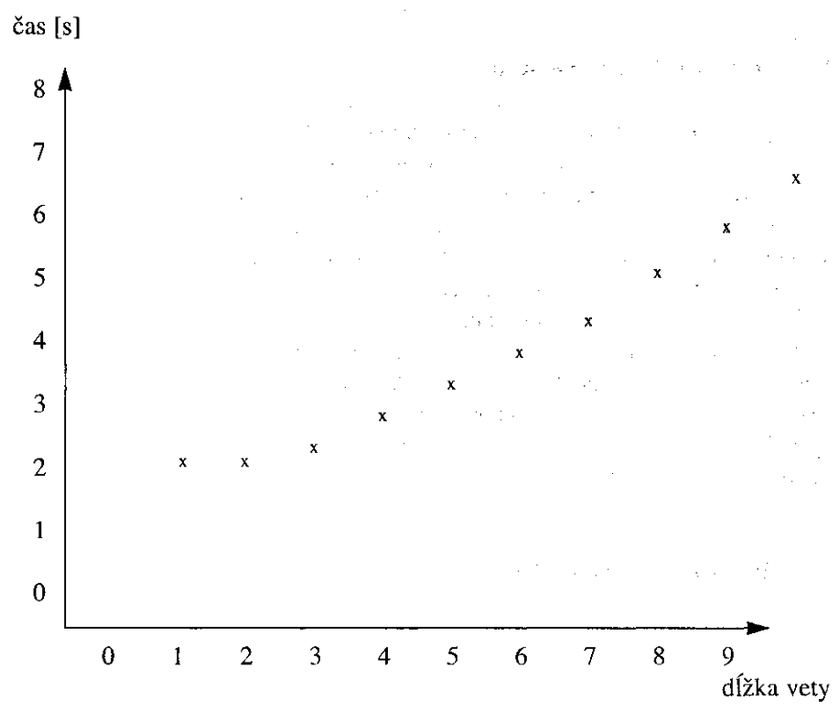
Veľkosť vstupu meriame *počtom plnovýznamových koreňov* vo vete, ktorý sa približne rovná jej dĺžke. Do dĺžky vety teda nepočítame predložky, zvratné *sa* apod.; na druhej strane kompozitá, ako napr. *drahokam*, počítame za dva členy (*drahý* + *kameň*), pretože pri syntaktickej analýze sa s nimi nakladá osobitne. Na počítači IBM PC AT 386DX, 40 MHz, 13 ms disk v interpretovanom ARITY-PROLOGu sme namerali tieto hodnoty:

počet vetných členov	1	2	3	4	5	6	7	8	9	10
priemerný čas [s]	2.2	2.2	2.5	3.0	3.4	4.0	4.5	5.2	5.9	6.8

Namerané hodnoty sa dajú aproximovať kvadratickou funkciou

$$T(x) = 0.05x^2 + 0.025x + 1.75$$

ktorá je však taká plytká, že sa podobá skôr funkcii lineárnej:



Uvážme, že priemerná desaťčlenná veta obsahuje už značný počet polysémických slov, morfologických a syntaktických viacznačností, ktorých kombinácie by mohli rásť do stoviek či tisícov.

Vidíme, že náš syntaktický analyzátor sa nespráva exponenciálne. Vďaka synergickému efektu postupuje bezmála deterministicky, takže čas výpočtu rastie zhruba len v priamej úmere k dĺžke vety. SAPFO teda pracuje v reálnom čase, a to je najlepším potvrdením našich teoretických úvah.

Pravda, treba povedať, že n = 10 je ešte príliš malé číslo na to, aby odstránilo obavy z kombinatorickej explózie, ktorá by sa mohla neskôr prejaviť pri vetách s dĺžkou 20, 30 a viac slov. Žiaľ, pamäťový priestor pre globálny zásobík ARITY-PROLOGu je obmedzený na 64 kB, takže pretečenie spravidla znemožní analyzovať vety dlhšie ako 10 slov.

Morfologická analýza jedného slova (spolu s vyvolaním celého slovotvorného hniezda a kompletnej k nemu prislúchajúcej informácie) vyžaduje konštantný čas v priemere cca 1.75 s, ktorý nezávisí od dĺžky vety.

Generovanie jednej parafrázy trvá v priemere 1.0 s. Samotných parafráz je však kombinatoricky veľký počet, takže problém sa posúva na selekciu medzi nimi.

17.

ETICKÉ IMPLIKÁCIE VÝSKUMU V UMELEJ INTELIGENCII

17.1 Cieľavedomé rozhodovanie v problémovom priestore

Dôležitou charakteristikou inteligentného subjektu je schopnosť predvídať, plánovať a robiť cieľavedomé rozhodnutia.

H. A. Simon (1983) definuje *rozhodovanie* (decision making) ako „heuristické hľadanie medzi alternatívami v problémovom priestore možných smerov konania s cieľom dosiahnuť preferovanú množinu následkov".

Najúspešnejšie programy v umelej inteligencii, ako STRIPS (Fikes - Nilsson, 1971) alebo SHRDLU (Winograd, 1972), boli založené na tomto princípe.

Reprezentatívnym príkladom rozhodovania v dobre definovanom problémovom priestore je šachová hra. V šachu sú *jednoznačne* definované objekty problémového prostredia, ich vlastnosti, vzájomné vzťahy ako aj cieľ hry. Prostredím je 64 polí šachovnice; množinou objektov je 32 figúr; každá má farbu a spôsob, ako s ňou možno ťahať; cieľom je dať súperovi mat. Počítač dokáže viac alebo menej exhaustívnym spôsobom prehľadať všetky možné postupnosti ťahov, ktoré môžu nasledovať, výsledné situácie ohodnotiť a vybrať najlepšiu z nich. Podľa toho potom ťahá.

Avšak reálne životné situácie sa vyznačujú práve tým, že problémový priestor nie je definovaný jasne. „Väčšina, ak nie všetky kľúčové rozhodnutia, s ktorými sú konfrontovaní manažéri, bývajú hmlisté problémy, ktorým celkom dobre nerozumejú ani oni sami, ani ich organizácia, a pri rozhodnutí sa stáva podstatný ich osobný úsudok" (Flores, 1982).

„Podstatou inteligencie je schopnosť konať primerane práve keď nie je daná žiadna jednoduchá pred-definícia problému alebo priestor stavov, v ktorom by sme vyhľadali riešenie. Racionálne prehľadávanie v problémovom priestore nie je možné skôr, než je tento priestor sám vytvorený, a je užitočné len do tej miery, do akej jeho formálne štruktúry efektívne zodpovedajú skutočnej situácii" (Winograd - Flores, 1987)[1].

Inteligentný subjekt teda musí predovšetkým *v záplave zmyslových vnemov rozpoznať to podstatné*. Musí najprv štruktúrovať prostredie do objektov, uvedomiť si ich podstatné vlastnosti a odhaliť zákonitosti, ktoré medzi nimi platia. No kontinuum nevyčerpateľne bohatej reality možno rozčleniť viacerými spôsobmi. Spôsob vyčlenenia podstatných prvkov a rozčlenenia problémového priestoru potom zásadne ovplyvní aj výsledok, ktorý nakoniec dostaneme.

---

[1] Počítače zatiaľ prejavili istú výkonnosť pri prehľadávaní programátorom preddefinovaného problémového priestoru, ale iba slabú, alebo vôbec žiadnu schopnosť problémový priestor samostatne utvárať, resp. rozpoznať, že starý problémový priestor už nie je v novej situácii adekvátny.

Problém vymedzenia (McCarthy - Hayes, 1969) stále zostáva otvorený. Aj keby počítač obsahoval *všetky* algoritmy a všetky pravidlá na všetky životné situácie, potreboval by ešte nejaký algoritmus,

17.2 Poznávanie s ohraničenou časovou a energetickou rezervou

Kognitívny subjekt má za úlohu poznávať zákon, čiže invariantné prvky v prostredí. Len spoľahnutie sa na trvale platné pravidlá mu umožňuje predvídať, plánovať a prežiť alebo dosiahnuť iný cieľ v jeho životnom prostredí. Avšak náš kognitívny subjekt[2] (človek, zviera, stroj) nielen že stojí pred touto úlohou, ale stojí pred ňou s ohraničenými časovými a energetickými rezervami. Je donútený poznávať prostredie a reagovať naň v ohraničenom čase a s minimálnymi energetickými stratami. *Mierou inteligencie je schopnosť primerane reagovať na čím väčšie zmeny prostredia za čím menší čas a s čím menšou energetickou stratou.*

Ak má byť reakcia subjektu ešte účinná, musí reagovať v konečnom čase. Ak má reagovať v konečnom čase, má možnosť vziať do úvahy iba konečné množstvo informácie. Poznanie každého kognitívneho subjektu bude vždy iba čiastočné a formulované iba na základe čiastkového súboru informácií.

Kognitívny subjekt je teda donútený robiť *selekciu*. *Musí* si všímať podstatné prvky a *musí* si *nevšímať* prvky nepodstatné. Ale na základe *čoho* rozhoduje, ktoré prvky sú pre neho podstatné viac a ktoré menej?

---

podľa ktorého by rozhodoval, ktoré z toľkých pravidiel je v danej situácii vhodné použiť - aby nepoužíval správne pravidlá na nesprávnom mieste a v nesprávnom čase.

Cudzie poznatky, ktoré sme získali tak, že sme ich prevzali hotové od druhých, sa zásadne líšia od poznatkov, ktoré sme získali sami. Ak som na nejaký poznatok prišiel sám, viem aj *na základe čoho*, na základe akej skúsenosti, informácie a v akej situácii som ho sformuloval. V zmenenej situácii si ľahko uvedomím, že poznatok nemám viac používať. *Len ten, kto poznatok sformuloval, je kompetentný ho zrušiť.*

Ak však prevezmem niečo cudzie z čírej „zbožnej" úcty pred autoritou a bez toho, aby som sám vedel, prežil, na základe čoho tento poznatok vznikol, som v takom „poznatku" chytený, uväznený; je to niečo, čo slobodnému ľudskému duchu iba zväzuje ruky. V takom prípade neviem ani kedy tento poznatok smiem použiť, ani kedy s jeho používaním už smiem prestať. To by musela prísť zase oná autorita, aby ma nasmerovala novým smerom.

Každé veľké zjavenie zo Svetla malo dve časti: vnútornú a vonkajšiu. Tá prvá je večná a nevýslovná, tá druhá je len premietnutím večnej pravdy do foriem - cez konkrétne historické, geografické a spoločenské podmienky v danom čase na danom mieste. Každé pravidlo a prikázanie bolo formulované s ohľadom na podmienky a s ohľadom na určitý cieľ.

Náboženské obce ešte po tisícročiach pokračujú v starých stereotypoch, hoci sa už geograficky ako aj po všetkých ostatných stránkach nachádzajú v odlišných podmienkach. Nedokázali samostatne rozlíšiť formu od obsahu; pretože v sebe nie sú živí, ale iba živé mŕtvoly; pretože svojmu náboženstvu nikdy nerozumeli a neprežili ho v živom cítení, ale majú ho iba naučené.

Slepé dodržiavanie pravidiel môže ísť tak ďaleko, že popiera aj samotný cieľ, kvôli ktorému bolo pravidlo pôvodne sformulované. Keď Mohamed prikázal nepiť alkohol, bolo to v záujme psychického a fyzického zdravia ľudí. No fanatický moslim radšej nechá brata chorého, ako by mu mal podať liek vylúhovaný v alkohole.

[2] Kognitívnym subjektom je aj biologický druh ako celok. Druh ako celok má tú výhodu, že sa môže množiť a reduplikovať na viac jedincov. Ako polysomatická (mnohotelová) bytosť si môže dovoliť aj experimenty končiace sa smrťou a poučiť sa z nich.

Jedinci biologických druhov, pretože majú obmedzenú výpočtovú kapacitu, využívajú prevažne zdedené vzorce správania a pri prírodných katastrofách, nečakaných udalostiach a náhlych zmenách prostredia zomierajú.

No biologický druh ako celok, resp. *život* ako taký, vďaka svojej mnohotelovosti a nepredstaviteľnej dlhovekosti mal obrovské časové a energetické rezervy a mal možnosť odhaľovať prírodné zákony v časovom rozpätí miliárd rokov a to ešte paralelne, experimentujúc na miliónoch jedincov súčasne.

Zadívajte sa na obrázok. Dal by sa nazvať „každý vidí, čo chce".

Čo vidíte? Mladú krásavicu, alebo škaredú starenu? Čo je na obrázku v skutočnosti?

Na obrázku sú čiary. Závisí od aktivity subjektu, ako tieto čiary pospája do celku, akú ideu, pojem im priradí.

Všimnite si, ako ťažko sa hľadá „mladá krásavica", ak sme už rozpoznali „škaredú starenu" alebo naopak. Rozum vytvorí ideu, na nej skostnateno lipne; oči znova a znova zbiehajú po tých istých líniách a my vidíme stále iba to, čo sme vopred mali vidieť.

Poznanie zabraňuje poznaniu, pravda?

17.3 Hodnoty štruktúrujú nediferencovanosť

Kognitívny subjekt musí rozčleniť kontinuum vnemov tak, aby toto rozčlenenie bolo účelné, funkčné vzhľadom na jeho cieľ. Musí si všímať tie vlastnosti objektov, ktoré sú v tej chvíli relevantné, podstatné vzhľadom na jeho hodnotový systém. A vzhľadom na rozličné ciele sú účelné rozličné rozčlenenia a rozličné prvky sa stávajú podstatnými.

Je to teda *hodnotový systém*, ktorý rozčleňuje *nediferencovanú skutočnosť* na objekty, ktoré potom nazývame „reálne existujúcimi"; a ktorý vyberá z *plnosti vnímania* tie prvky, ktoré budeme vnímať a tie, ktoré vnímať nebudeme.[3]

Na tento *podvedomý predvýber faktov* upozornil H. Poincaré (1913). Ukázal, že ešte aj v základoch takej exaktnej vedy, akou je fyzika, stoja etické a estetické hodnoty, ako krása a pod. Z množstva hypotéz, ktoré sa ponúkajú na vysvetlenie empirických dát, fyzik bez výčitiek svedomia vyberá jednu a nehanbí sa povedať, že je pravdivejšia, pretože je jednoduchšia, elegantnejšia, účelnejšia.

T. S. Kuhn (1981) ukázal, že veda robí takýto predvýber už pri voľbe empirických dát, ktoré hodlá alebo nehodlá spracovávať. Postupuje teda nevedecky, pretože si vyberá iba niektoré fakty s cieľom podporiť a zúplniť už vopred danú paradigmu myslenia. Ostatné ignoruje. Až keď sa tlak faktov na *nevnímanej strane reality* nahromadí, vynúti si kolaps starej paradigmy. Vznik novej paradigmy sprevádza definovanie nového pojmového aparátu, ktorý tú istú realitu štruktúruje úplne iným spôsobom.[4,5]

---

[3] Porovnaj výroky z Bhagavad-Gíty a indických sútier. „Len (emocionálnym) ľpením vzniká svet (objektov pre nás)." Odpútanosť je nevyhnutná pre nesprostredkované, čisté vnímanie. (Radhakrishnan, 1958; Pataňjali, 1988).

[4] R. P. Feynman (1990) je toho názoru, že *skutočný* pokrok v matematike, fyzike (a vo vede všeobecne) nespočíva v dokazovaní nových viet, ale práve v premene a dozrievaní pojmového aparátu, ktorý nám umožňuje uchopiť realitu novým a oveľa lepším spôsobom.

Ako príklad podvedomej selekcie nám môže poslúžiť skúsenosť, ktorú iste všetci poznáme. Pri vnímaní ľudskej tváre je dôležité, aby sme dokázali rozoznať osoby medzi sebou. Každý Európan má skúsenosť, že príslušníci čiernej a žltej rasy mu pripadajú všetci rovnakí. To isté však hovoria černosi a Ázijci o príslušníkoch bielej rasy! Človek dobre rozoznáva tváre svojich súkmeňovcov, inak má ťažkosti.

Psychológovia vysvetľujú tento jav tým, že na ľudskej tvári sú určité podstatné prvky, ktoré sú na rozlíšenie tvárí medzi sebou dôležité, dištinktívne. Tieto dištinktívne prvky sú však u každej rasy iné. Európan sa sústreďuje na prvky, ktoré sú dištinktívne pre Európanov, avšak tie môžu byť u iných rás u každého príslušníka práve rovnaké.

Sústredením vzniká určitý druh *slepoty* pri vnímaní. Zaujímavé je, že túto slepotu, nevhodnosť starých konceptov, sme spravidla schopní odhaliť až potom, keď nastane nejaké zlyhanie v praktickom živote, v aplikačnej oblasti.

Iný príklad. Jeden chemický inžinier vstúpil do archeologického múzea a v jednom náleze označenom ako 'kultový predmet' rozpoznal Voltov elektrický článok. Archeológovia, ktorí sú špecialisti vo svojom odbore, nie sú predsa chemici a nemôžu vedieť, kedy sa niečo podobá na Voltov elektrický článok. A tak pri mnohých nálezoch stála nálepka 'kultový predmet'. Paradoxné je, že *špecializácia archeológov v archeológii ich urobila zlými špecialistami v ich vlastnom odbore!*[6]

Africké domorodé kmene, žijúce na savane, majú vyše sto výrazov pre rôzne druhy tráv, Eskimáci práve toľko výrazov pre rôzne druhy snehu, lesní obyvatelia tisíce výrazov pre rozličné dreviny a púštni beduíni *dvesto* pomenovaní pre ťavu! Keď Európan príde na savanu, vidí na mieste, kde domorodec vidí celý rad rozlič-

---

[5] Zdá sa, že štruktúra vedeckých revolúcií (v zmysle T. S. Kuhna, 1981) sa v niečom podobá na štruktúru politických revolúcií. Exponenti starej paradigmy, ktorí celý čas potláčali nové, sa vo chvíli svojej porážky chcú postaviť na čelo a sami viesť spoločnosť novým smerom.

Aj vo vede sa všetko radikálne nové môže formovať iba v rámci tzv. alternatívnych vied, ktoré sú v akomsi disidentskom postavení. Alternatívna medicína (homeopatia, akupunktúra) bola ešte v minulých desaťročiach terčom neľútostného výsmechu lekárov, hoci teraz sú to práve oni, ktorí ju chcú robiť vedecky, a najradšej by z nej vylúčili tých, ktorí ich do nej zasvätili. V súčasnosti takým istým procesom prechádza alternatívna psychológia (transpersonálna psychológia, psychotronika). A *celý rad* vedných oblastí na svoje zverejnenie a využitie ešte len čaká, pretože väčšine ľudstva ešte pýcha a zúžené vedomie zabraňujú uvedomiť si niečo nové.

Seriózne vedecké experimenty (Sheldrake, 1989; Grof, 1985) sú nejedným vedcom stále bezmyšlienkovite popierané bez toho, že by sa unúval informovať sa o nich. Svetová akademická veda je kupodivu v zajatí akéhosi naivno-materialistického predsudku, takže sa vyhýba diskusii o všetkých jemnejších svetoch a bytostiach bez afektu a na báze zdravého rozumu; hoci tieto sú taktiež materiálne, a tiež podliehajú exaktným fyzikálnym zákonom. Tieto zákony sa dajú dokázať opakovateľnými merateľnými experimentami a mnohé z nich sú známe, žiaľ, nie našej vede, ktorá by vlastne mala stáť na čele výskumu, namiesto toho, aby ho nepriamo hatila prenechávajúc voľné pole pôsobnosti najrozličnejším šarlatánom a eskamotérom.

[6] Porovnaj výroky Bhagwana Radžníša: „Obmedzenie vedomia koncentráciou je existenčná nevyhnutnosť. Za to nie je nikto zodpovedný. Pri tomto obmedzení sa však nesmie zostať. Inak sa všetky možnosti vedomia strácajú" (Rajneesh, 1982).

Myslenie je prehľadávanie poznaného. Meditácia je vyslobodenie z poznaného... „... je to nábeh ku skoku do neznáma".

Meditácia je proces inverzný k mysleniu. Myslenie je proces zaostrenia, zúženia vedomia. Modlitba a meditácia znamenajú návrat k celistvosti, zrieknutie sa čiže rozšírenie vedomia.

ných bylín, jednoducho iba „trávu". Tu nejde iba o rozdiel v pomenovaní, ale o dve skutočne odlišné vnímania.[7,8] Toto rozličné vnímanie sa potom zreteľne prejaví v konaní.

Európan bude vnímať flóru na savane len ako „trávu" dovtedy, kým raz jednu z tých tráv zje a priotrávi sa. Od tejto skúsenosti sa mu predtým nediferencovaný pojem „trávy" rozčlení na dva, ktoré sa mu navždy vryjú do pamäti. Všimne si, že tu existujú prinajmenšom dva druhy trávy, a zapamätá si nejaký dištinktívny príznak medzi nimi. Teda objekty sveta vystúpia z nediferencovanosti a začnú existovať pre nás až vtedy, keď majú vzťah k niečomu, čo je pre nás dôležité.

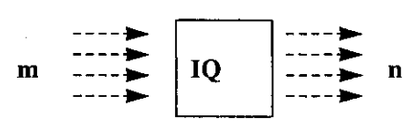
Na tento fakt upozornil aj M. Heidegger (1968). Podľa Heideggera veci začínajú pre nás existovať až vo chvíli nejakého zlyhania, nedostatku (breakdown) - inak splývajú s pozadím. Všetci vieme, že človek začína vnímať svoje zdravie až vtedy, keď ho nemá. Jednotlivé telesné orgány si neuvedomujeme dovtedy, kým fungujú bezporuchovo. Až keď niektorý z nich zlyhá, uvedomíme si jeho existenciu veľmi zreteľne.

Ako zhrnutie tohto odseku môžeme povedať, že *spôsob, akým si človek utvára svet okolo seba, je premietnutím jeho hodnotového systému.* To platí tak o vnímaní, ako aj o konaní.

Špeciálne pri rozhodovaní si človek problémový priestor utvára účelne vo vzťahu k svojim cieľom a vníma iba tie prvky, ktoré sú relevantné z hľadiska jeho chcenia. Ako by mohol vyzerať algoritmus, ktorým by si kognitívny subjekt svoj problémový priestor pre rozhodovanie dokázal samostatne tvoriť?

17.4 Model umelej inteligencie

Abstrahujme inteligentný subjekt ako m x n - rozmernú funkciu, ktorá zobrazuje m-ticu vstupov na n-ticu výstupov v prostredí, v ktorom má za úlohu prežiť, resp. splniť iný želaný cieľ.



Upozorňujeme, že m,n sú v praxi veľké čísla (10^5,10^6 a viac). Ako vstupy (n) si predstavte bunky sietnice oka, ucha, tepelné, ťahové, tlakové senzory pokožky;

[7] H. R. Maturana (1970) skúmal proces vizuálneho vnímania u žiab a dospel k takémuto záveru: „Čoskoro som si vo svojom hľadaní uvedomil, že predmetom mojej štúdie farebného vnímania nie je štúdia o zobrazení farebného sveta do nervového systému, ale že ide skôr o snahu porozumieť, akým spôsobom sa sietnica (a nervový systém) podieľa na *utváraní* farebného priestoru pozorovateľa."

[8] B. L. Whorf (1956) správne upozorňuje, že už samotný pojmový aparát jazyka definuje problémový priestor tak, že sugestívne naznačuje určité spôsoby vnímania a správania.

alebo CCD kamery rôznych spektrálnych dĺžok, zvukový snímač a pod. Výstupy (m) sú motorické (nejaký druh pohybového ústrojenstva a rúk), ale aj iné signály, reč atď.

Prostredie pozostáva z pestrej škály javov, správajúcich sa podľa zákona (množiny pravidiel), ktoré sú však subjektu neznáme.

Akákoľvek úspešná činnosť subjektu v prostredí predpokladá *schopnosť predvídania* chodu udalostí, schopnosť urobiť všetko, čo treba, dostatočne včas.

Schopnosť predvídania a adekvátneho reagovania závisí od miery poznania zákona, podľa ktorého sa javy v okolitom prostredí správajú.

Prostredie nie je dokonale entropické, ale vykazuje štruktúrovanosť, opakovanie, pravidelnosti, zákonitosti. Subjekt bude postupovať hľadaním invariantov (večnej pravdy), t. j. bude si všímať všetko, čo sa opakuje, čo sa vyskytuje vždy spolu (v čase alebo v priestore). Získané hypotézy o invariantoch si sformuluje do pravidiel, nazve ich zákonmi a spolieha sa na ne pri predvídaní budúcnosti.

Prírodu uvažujme ako postupnosť totálnych stavov s_1, s_2, s_3... v časových okamihoch t_1, t_2, t_3 ... danú rekurentne:

$$s_0,\ s_1 = F(s_0),\ s_2 = F(s_1)\ \ldots\ s_k = F(s_{k-1})\ \ldots$$

F je funkcia, zákon, ktorý na základe predchádzajúceho okamihu definuje spôsob pohybu (vývoja) v prírode v okamihu nasledujúcom. Prírodu sme zjednodušili na mechanisticko-deterministický stroj pracujúci v diskrétnom čase, čo bude pre naše ciele irelevantné.

Ak predpokladáme, že funkcia F, podľa ktorej sa prostredie správa, je obmedzená na nejakú špeciálnu matematickú triedu (lineárne, kvadratické, polynomiálne funkcie s n stupňami voľnosti), je úloha umelej inteligencie riešiteľná v polynomiálnom čase.

Ak však pripustíme, že v prostredí fungujú nielen polynomiálne, ale aj ľubovoľné iné závislosti, potom je úloha umelej inteligencie NP-ťažká a jej výpočtová (časovo-energetická) zložitosť je superexponenciálna.

Ak pripustíme, že ľubovoľné dva prvky prostredia na vstupe môžu spolu súvisieť, potom počet možných funkcií F je kombinatoricky exponenciálny a aj učiaci algoritmus musí tomu zodpovedať.

Toto naznačujú všetky doterajšie pokusy o strojové samoučenie. Všetky programy, ktoré fungovali v reálnom čase, predpokladali štrúkturované prostredie s nízkou entropiou (mierou neusporiadanosti). Funkciu prostredia potom hľadali v zjednodušenom tvare (Winston, 1975, 1977). Napríklad perceptrón v podstate zodpovedá lineárnej klasterizácii, t. j. predpokladá, že objekty vo svete sú lineárne oddeliteľné a dokáže odhaľovať iba lokálne súvislosti medzi nimi (Minsky - Papert, 1969).

Všetky tie programy, ktoré nekládli ohraničenia na hľadanú funkciu, zlyhali na superexponencialite výpočtovej zložitosti.

Pre náš výpočtový subjekt by totiž teoreticky existovalo jedno triviálne riešenie: prehľadať všetky kombinácie svojich vstupov vo všetkých časových okamihoch

v pamäti, porovnať ich medzi sebou a izolovať exhaustívnou metódou spoločný prvok, invariant. Exhaustívny algoritmus by vyžadoval milión na milióntu operácií alebo podobné ešte vyššie čísla. Teda prostredie je poznateľné, ale v nereálnom čase. To by odporovalo definícii inteligencie, pretože inteligentný subjekt musí byť schopný reagovať na pohyby okolo seba včas, t. j. skôr, ako sa stane niečo zlé.

Ak subjekt obmedzí tvar, v ktorom hľadá funkciu F, na určitú zúženú triedu funkcií (závislostí), je schopný odhaľovať jednoduchšie zákonitosti a sformulovať si čiastočné poznanie v reálnom čase. To potom využíva. Vtedy však môže iba dúfať, aby do hry nevstúpila niektorá z tých zložitostne vyšších zákonitostí, ktorých si nie je vedomý.

17.5 Vznik automatizovaných vzorcov správania

Inteligencia, ktorá má k dispozícii ohraničenú výpočtovú a pamäťovú kapacitu, si v súboji s časovou tiesňou vytvorila zefektívňujúci mechanizmus: je to *zúženie vedomia* alebo *automatizovaný vzorec správania*.

Inteligencia - buď vďaka tisícročnej skúsenosti, alebo nesmiernemu intelektuálnemu úsiliu v čase, keď práve nie je pod existenčným tlakom - odhalí nejaký zákon sveta. V oboch prípadoch to urobí za cenu nepredstaviteľného prehľadávania a veľkej energie, po nespočetnom množstve pokusov a omylov.

Potom sa snaží sformulovať svoj objav do '*minimálneho vyjadrenia*.' To urobí tak, že oddelí relevantné znaky (vstupy a ich hodnoty) od nerelevantných. Na rozpoznanie nejakej štandardnej situácie (frame, pattern) môže byť (a vďaka malej entropii sveta obyčajne aj je) rozhodujúcich len niekoľko špecifických znakov, hodnôt na niekoľkých málo z bezpočtu vstupných kanálov. Na ostatných potom pri rozpoznávaní tejto situácie nezáleží. Pri procese identifikácie štandardných situácií (pattern matchingu) ich procesor môže *„zabudnúť"*, *„zahmliť"*, *„zakryť závojom"*, jednoducho vylúčiť zo svojej pozornosti, z poľa svojho vnímania. To je *zúženie vedomia*. Umožňuje ušetriť čas a energiu a teda prežiť.

Príklad. Náš IQ-automat zistil, že pre identifikáciu predmetu (situácie) XYZ je rozhodujúce a postačujúce, aby na vstupe č. 11 a 27 bola vnímaná červená farba, na vstupe 101 zelená a na č. 1056 zvuková vibrácia frekvenciou 90 Hz.

Táto vstupná konfigurácia spoľahlivo identifikuje objekt XYZ, *lebo* (!) počas procesu učenia (prehľadávania veľkého množstva možností) sme sa presvedčili, že v prostredí, v ktorom sa pohybujeme, neexistujú už žiadne iné objekty dávajúce na vstupoch 11, 27, 101, 1056 tieto isté hodnoty. Teda štvorica (trigger configuration)

XYZ	
11:	červená
27:	červená
101:	zelená
1056:	90 Hz

je pre XYZ *charakteristická*. Je zároveň aj *minimálna*. To znamená, že ak by sme vynechali ktorúkoľvek položku z rámca, mohlo by dôjsť k zámene s inými objektmi. Minimálnosť je tu zase *vďaka tomu* (!), že počas procesu učenia sme sa presvedčili, že v prostredí sa vyskytujú aj iné objekty, napríklad KLM, ktorý sa prejaví na vstupe

KLM	
11:	modrá
27:	červená
101:	zelená
1056:	90 Hz

Teda práve '11:modrá' je dištinktívny príznak, ktorý odlišuje XYZ od KLM a treba si ho všímať.

Odteraz si naša abstraktná inteligencia môže dovoliť reagovať na XYZ alebo KLM v lineárnom čase. To jej umožňuje reagovať s maximálnou rýchlosťou a minimálnymi energetickými nákladmi.

Ak sa napríklad na vstupe objaví '11: červená', stačí už len overiť, či aj '27: červená', '101: zelená' a '1056: 90 Hz'. Ak áno, nepochybne ide o XYZ. Kontrolou ostatných vstupov netreba strácať čas. Opakovať v časovej tiesni vždy proces učenia, alebo sa zaoberať inými vstupmi, v čase keď pozornosť musí byť sústredená na ohnisko diania, je nemysliteľné.

Takéto automatické vzorce reagovania na prostredie sa fixujú u zvierat ako inštinkty a u ľudí navyše aj ako návyky, na vyššej úrovni honosne nazývané „poznatky".

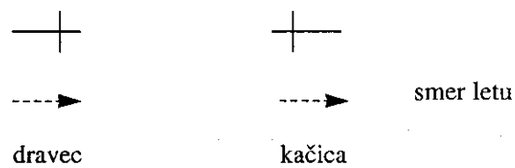
Čo však nastane, ak sa prostredie, objekty v prostredí a ich vlastnosti náhle zmenia?

17.6 Vrodené spúšťacie schémy u zvierat

„Podnety vyvolávajúce inštinktívnu reakciu ako prvý opísal jeden zo zakladateľov etológie a nositeľ Nobelovej ceny z roku 1973 rakúsky vedec *Konrad Lorenz* (1943). Špecifické podnetové znaky, prvky vrodených spúšťacích schém, ktoré jediné priamo vyvolávajú vrodenú odpoveď živočícha, nazval Lorenz *kľúčové podnety*. V pokusoch využil umelé atrapy, ktoré zdôrazňovali podstatné tvary alebo sfarbenie určitej časti tela a vyvolávali vrodené reakcie. V tejto súvislosti možno spomenúť pokus z roku 1935 pri výskume obranného správania kačíc, husí a kurovitých vtákov (napr. bažanty, kury, prepelice).

K. Lorenza zaujal problém, ktorý znak na tele je najvýznamnejší pri vyvolávaní obranného správania pri prelete dravca nad pokusným poľom. Pri náhodnom prelete živého dravca (myšiaka hôrneho, sokola sťahovavého alebo orla skalného) prejavovali kurovité vtáky obranné, strachové a únikové reakcie a okamžite vyhľadali úkryt. Keď bádateľ ťahal na lane atrapu dravca nad pokusným poľom chvostom

dopredu, úteková reakcia sa neprejavila, ale atrapa, ktorá sa pohybovala dopredu hlavou, teda v prirodzenom letovom obraze, okamžite vyvolala útekovú reakciu. Letový obraz dravcov je *„krátkokrký"* a *„dlhochvostý"*. Letový obraz kačíc *„dlhokrký"* a *„krátkochvostý"*:



Tak sa zistilo, že na vyvolanie útekovej reakcie je potrebný ako kľúčový podnet jednak smer letu dravca a jednak jeho letový obraz s najpodstatnejšou časťou - krátkym krkom. Vrodená spúšťacia schéma uvedie do chodu tzv. *fixný motorický prejav*, v tomto prípade útekovú reakciu" (Novacký - Czako, 1987).

Kačice, ktoré svoj mechanizmus reakcie na dravce vyvíjali počas miliónročí vo voľnej prírode, nepočítali s tým, že sa v prostredí odrazu objaví úplne nový objekt - Konrad Lorenz, a že im bude nad hlavami ťahať papierové atrapy. Kačica reaguje výlučne na základe dvoch relevantných znakov: krížovej siluety a jej orientácie. Ostatné je irelevantné, nezaujímavé, nevnímané.

Preto označíme kačice ako „sprosté", lebo sa nechajú plašiť atrapami. A to je ešte prípad, keď im to prejde ľahko. Je bezpočet iných prípadov, keď zvieratá hynú novosaobjavivšími činnosťami človeka, na ktoré nemajú vypracované rozpoznávacie a výkonné mechanizmy. Zahynú nie preto, že by nemali prostriedky utiecť alebo sa brániť, ale preto, že situáciu vôbec nerozpoznávajú a nechajú sa podrezať ako teľce.

17.7 Naučené spúšťacie schémy u človeka

U človeka sa bežne uplatňuje ten istý mechanizmus. Rozdiel je v tom, že u človeka prevažujú nad vrodenými spúšťacími schémami *schémy získané vzdelaním.*

Nie že by človek, ako duchom obdarená bytosť, nemal potenciálnu schopnosť sebareflexie, vyslobodenia z podmienenosti vlastných stereotypných reakcií; tragédia je v tom, že on tento dar vo svojej nekonečnej duchovnej lenivosti nehodlá používať, kým mu nejde o kožu; aj potom len zriedka, ak sa vôbec na svoju zabudnutú a dávno zakrnelú schopnosť dokáže rozpamätať.

V najprimitívnejších prípadoch sa zúženie vedomia prejavuje napr. posudzovaním ľudí podľa obmedzeného počtu príznakov, ktorých počet sa môže redukovať až na jeden: napr. farba pleti, pohlavie, vek, akademický titul, príslušnosť k špecifickej skupine, národu, vierovyznaniu, politickej strane alebo výška valutového konta, obvod poprsia a iné.

V ďalších prípadoch ide o komplex zložitejších ale inak rovnako mechanických spúšťacích mechanizmov.

V najúctyhodnejšom prípade ide o „súbor poznatkov a presvedčení" tzv. inteligenta, ktorý je však rovnako nemenný, rigídny a mechanický a robí ho prakticky neosloviteľným a kooperácie neschopným, akonáhle ide o niečo nové.

Záver: *Myslenie bez sebareflexie je jeden veľký súbor predsudkov. Kultivovanie poznania bez sebapoznania vedie k ignorancii.*

17.8 Kreativita nad racionalitu alebo srdce nad rozum

Rozdiel medzi *živým duchom* a *mechanickým rozumom* ilustruje vtip: Máte na výber medzi dvojo hodinkami. Jedny stoja. Druhé sú také, že idú stále päť minút popredu. Ktoré si vyberiete?

Počítač: Prvé hodinky sú dokázateľne lepšie, lebo ukazujú presný čas aspoň dvakrát denne, zatiaľ čo tie druhé nikdy.

Človek: Vyberiem si tie druhé hodinky. Päť minút si vždy odrátam.

Rozdiel medzi týmito odpoveďami je, že prvá je síce *logická*, ale druhá je *kreatívna*.

17.9 Zúženie a rozšírenie vedomia

Proces učenia (tvorenia) je nepomerne časovo a energeticky náročnejší ako proces používania (opakovania) už overeného vzorca konania. Podľa toho potom inteligentný subjekt funguje vo dvoch režimoch:

(a) režim učenia (tvorivosti)
(b) režim (mechanického) používania naučeného

Príroda doposiaľ využívala *(a)* len vo vývoji druhu ako celku.[9] V správaní jedincov silne prevažovalo *(b)* nad *(a)*.

Človek je jediná pozemská bytosť, ktorá má kompetenciu na vedomé *(b)* aj *(a)* v plnej miere.

Avšak to, čo sa zatiaľ všeobecne nazýva *myslením*, je len prehľadávanie pamäte na staré poznatky, t. j. čisté *(b)*. A to, čo sa nazýva *učením*, pre väčšinu znamená len preberanie hotových vzorcov od druhých ľudí a z tradície, čiže zase len čisté *(b)*. Všetko nové je tu vďaka niekoľkým inšpirovaným jednotlivcom.

[9] Minerálne, rastlinné a živočísne druhy (nie jedinci) sú skutočne kognitívnymi subjektami v plnom zmysle slova. Na rozdiel od človeka, ktorý má individuálnu seba-vedomú dušu, zvieratá nie sú si vedomé seba ako jedinci, ale majú vedomú tzv. kolektívnu dušu (Abd-ru-shin, 1990), inak zvanú aj dévická inteligencia (Mac Lean, 1980) alebo morfogenetické čiže tvaro-tvorné pole (Sheldrake, 1988). Toto jemnohmotné a vysoko senzitívne pole je spoluzodpovedné za morfogenézu ľudského plodu, ako aj genézu všetkých zvieracích, rastlinných a minerálnych foriem a je zároveň zhromaždiskom skúseností druhu ako celku.

Na prechode medzi *(a)* a *(b)* pozorujeme dvojsmerný proces:

I. Prechod od *(a)* k *(b)*. Samostatné učenie a fixácia naučeného. Zúženie vedomia.

II. Prechod od *(b)* k *(a)*. Odučenie. Rozšírenie vedomia. Inverzné k *I.*

Duchovne neprebudený subjekt trávi život tým, že hromadí a fixuje poznávacie vzorce - v lepšom prípade samostatne, v horšom ich iba preberá a opakuje. V oboch prípadoch však prejavuje výrazne zníženú schopnosť odpútať sa od svojich intelektuálnych schém, zostáva v nich chytený, obmedzený a zrieka sa ich iba pôsobením silného vonkajšieho tlaku a trápenia.

Polointeligenta (alebo počítač), ktorý má v danom prostredí naučenú tú správnu množinu 'kľúčových vzorov' a 'odpovedí', nemusíte dlhý čas rozpoznať. Bezpečne ho spoznáte až podľa jeho neschopnosti vykonať spätný proces *II*, t. j. odnaučiť sa starý vzor a vykonať odznova proces *I*, čiže samostatne utvoriť nové vzorce v novom prostredí. Taký subjekt sa prejaví tým, že v nových podmienkach naďalej reaguje podľa starých pravidiel. Teda *strojovosťou* typickou pre počítač, ktorého program bol vyrobený pre situáciu, ktorá sa medzitým zmenila.[10]

Naopak duchovne bdejúci subjekt je pripravený k obojsmernému procesu v každej chvíli - *reaguje na nové po novom*, čo sa nazýva *odpútanosťou* alebo *živosťou ducha*.[11]

To znamená, že ak napríklad netopier zomrie v osamelej jaskyni na ostrove a nezanechá po sebe žiadne potomstvo, jeho skúsenosť nevychádza nazmar. Zostane odpečatená v morfogenetickom poli netopierov a pri určitom kritickom množstve takýchto skúseností začína ovplyvňovať správanie druhu ako celku. Informácie medzi minerálmi, rastlinami a zvieratami sa dokazáteľne prenášajú bez ohľadu na hrubohmotný čas, priestor a bez akejkoľvek genetickej súvislosti.

[10] Mechanicky sa teda nesprávajú iba stroje. Mechanicky reagujú ľudia, mechanicky a bezcitne fungujú úrady.

[11] Správanie prevažnej väčšiny jedincov ľudského rodu po tisícročia bolo a doteraz stále je *stádovité*, t. j. obmedzuje sa na preberanie a používanie vzorcov správania od tzv. *tvorivých jedincov*.

Takáto deľba práce je na mieste, keď vedci preberajú hotové výsledky od svojich kolegov; ale stáva sa pohromou, ak sa aplikuje na oblasti, v ktorých sa každý musí rozhodnúť sám.

V pozadí všetkých fašizmov, komunizmov, rasizmov, šovinizmov, fundamentalizmov a všetkých ostatných *izmov* stojí práve oná neschopnosť *samostatne* rozlišovať, nedostatok tvorivej flexibility a žalostné zúženie vedomia.

V minulosti sa nešťastiu predchádzalo inštitúciou *úcty k vyšším stavom*. Šľachta a duchovenstvo mali byť *okom* národa a tí, čo boli odkázaní na opakovanie po druhých aspoň vedeli, ku komu sa majú obrátiť. Ale v storočí, keď sme už podkopali tradičné piliere mravnosti, ale väčšina národa ešte nie je kompetentná myslieť samostatne, nech nám nebesá pomáhajú.

Súčasná demokracia je založená na nepravdivom princípe; a na Aténsku demokraciu sa už vôbec nepodobá. Princíp, že *väčšina má pravdu*, je očividne nekorektný a ignoruje evidentný fakt, že jeden vidiaci je platný viac ako zástupy nevidiacich. Taká spoločnosť potom sotva môže vyzerať ináč, ako keď sa slepí snažia viesť slepých.

Pravda nemá nič spoločné s kvantitou ale s kvalitou. Je viac ako isté, že tí najmúdrejší a najušľachtilejší budú v menšine oproti masám, ktoré sú priemerné. Každá novovznikajúca pravda dokonca *musí* byť spočiatku v menšine. Bezcharakterná demokracia, ktorá nepozná žiadne hodnoty a propaguje čiste počítateľnú kvantitu bez akýchkoľvek kritérií na kvalitu, je len nástrojom na preferovanie všetkého priemerného až podpriemerného a na zahubenie všetkého najušľachtilejšieho a najcennejšieho.

17.10 Globálna kríza ľudstva

Z vyššie uvedených dôvodov ani riešenie globálnych problémov ľudstva neleží v oblasti, v ktorej sa ho väčšina ľudstva stále ešte snaží hľadať. *Riešenie globálnej krízy nie je a nebude možné nájsť žiadnym druhom racionálneho hľadania samým osebe*. Samotný problémový priestor, v ktorom hľadáme, je zle definovaný a riešenie sa v ňom nenachádza, pretože je premietnutím pokriveného systému hodnôt.[12] Tentokrát už ľuďom nepomôže žiadna chytrosť, žiadna ešte dômyselnejšia rekombinácia súčasných spoločenských štruktúr ani výmena vládnucej strany - ak si všetci budeme chcieť ponechať svoje slabosti. *Nevyhnutná je revolúcia srdca!* Ľudia budú *musieť* vrátiť cnostiam ich staré miesta, ktoré im prináležia.[13]

Jediný fakt, na ktorom sa dokázali zhodnúť účastníci svetových konferencií o globálnych problémoch ľudstva ([1], [2]) je tento: za súčasných ekologických a spoločensko-ekonomických podmienok a vzťahov, ktoré sú premietnutím dnes dominujúcich etických hodnôt, neexistuje riešenie. Záchrana je možná iba za predpokladu, že ľudstvo bude ochotné urobiť presuny vo svojom rebríčku hodnôt. Ľudstvo bude musieť postaviť svoj hodnotový systém znova z hlavy na nohy: s duchovnými hodnotami hore a materiálnymi dole.

Globálne podmienky sa zmenili, a to celkovou technologickou a duševnou zrelosťou ľudstva, ako aj preto, že Modrá planéta vstupuje do vyžarovacieho poľa iných kozmických síl (Bailey, 1947, 1957; Abd-ru-shin, 1990). Staré vzory správania už nie sú adekvátne v novej situácii a vedú ku kolapsovým stavom.

Ak ľudstvo nebude chcieť vystúpiť zo svojej slepoty a nebude ochotné používať svoje duchovné schopnosti dobrovoľne, budú to hromadiace sa globálne katastrofy, ktoré ho k rozšíreniu vedomia donútia nedobrovoľne a bolestivo.[14]

A nie je ani pravda, že by neexistovali kritériá, ako rozoznávať kvalitu a človeka, ktorý je úcty hodný a ktorý by nás mohol viesť od toho, kto nie. Ale to by sme museli vedieť používať aj naše duchovné dary a naše srdcia, nie vždy len mechanický rozum!

[12] Dnes viac ako inokedy by sa svet mal zamyslieť, skôr než bude riešiť „dôležité" otázky bytia, či sú tieto otázky vôbec zmysluplne formulované. Skôr ako vynaloží miliardy na dosahovanie nejakých cieľov, mal by sa najprv pýtať, vzhľadom na aké hodnoty by tieto ciele mali byť žiadúce.

[13] G. K. Chesterton (1993) vtipne poznamenal, že *blázon nie je ten, kto stratil rozum, ale ten, kto stratil všetko okrem rozumu.* Svet blázna nie je nelogický. Svet blázna je až prekvapivo logický, uzavretý a rozumný; všetko je v ňom dobre zdôvodnené. Len oproti realite je akosi zúžený, ochudobnený a sploštený len na niekoľko málo dimenzií.

[14] Čínsky znak pre slovo 'kríza' sa skladá z dvoch znakov: 'nebezpečie-šanca'. V tomto období kvalitatívneho zlomu je ľudstvo vystavené tlakom, ktoré preň znamenajú veľké nebezpečie, ale aj šancu na nebývalý vnútorný rast.

Pramene

ABD-RU-SHIN: Im Lichte der Wahrheit. Gralsbotschaft. Stuttgart, Stiftung Gralsbotschaft 1990. 1091 s.

ALLEN, J.: Natural Language Understanding. Menlo Park, Benjamin-Cummings 1987. 537 s.

ALTMANN, G.: Wiederholungen in Texten. Bochum, Brockmeyer 1988. 230 s.

ALTMANN, G. - KÖHLER R.: Synergetické aspekty jazykovedy. Jazykovedný časopis 42, 1991, s. 3-15.

APRESIAN, J. D.: Lexical Semantics. Synonymical Means of Language. Moskva, Nauka 1974. 368 s.

AUSTIN, J. L.: How to do Things with Words. New York, University Press 1962.

BAILEY, A. A.: The Destiny of Nations. Treatise on the Seven Rays. Kent, Lucis Press 1947. 220 s.

BAILEY, A. A.: The Externalization of the Hierarchy. Treatise on the Seven Rays. Kent, Lucis Press 1957.

BALLY, CH.: Traité de stylistique francaise (rus. prekl. Francuzskaja stilistika). Moskva, 1961.

BARTON, G. E. Jr. - BERWICK, R. C. - RISTAD E.S.: Computational Complexity and Natural Language. Cambridge, MIT Press 1987. 335s.

BERWICK, R. C.: The Acquisition of Syntactic Knowledge. Cambridge, MIT Press 1985. 368s.

BIRKHOFF, G. - MCLANE, S.: A Survey of Modern Algebra. New York, MacMillan 1965. 480 s.

BOBROW, D. G. - KAPLAN, R. - NORMAN, D.: - THOMPSON, H. - WINOGRAD, T.: GUS - a frame-driven dialog system. Artificial Intelligence, 8, 1977, s. 155-173.

BOSÁK, J. - BUZÁSSYOVÁ, K.: Východiská morfémovej analýzy. Morfematika. Slovotvorba, 1985, s. 101-104.

BRANICKÁ, J.: Ustálené adjektívne prirovnania. Jazykovedný časopis, 42, 1991, s. 115-121.

BRODOWSKA-HONOWSKA, M.: Zarys klasyfikacji polskich derywatów. Wroclaw, Ossolineum 1967. 100s.

BUZÁSSYOVÁ, K.: Sémantická štruktúra slovenských deverbatív. Bratislava, Veda 1974. 236 s.

CAPELLI, A. - FERRARI, G. - MORETTI, L. - PRODANOF, I.: A Framework for Integrating Syntax and Semantics. Pisa, 1985. 1 s.

CAPRA, F.: The Turning Point. New York, Bantam 1983. 464 s.

CARPENTER, B.: The Generative Power of Categorial Grammars and Head-Driven Phrase Structure Grammars with Lexical Rules. Computational Linguistics, 17, 1991, 3, s. 301-313.

CLIFFORD, J.: Natural Language Querying of Historical Databases. Computational Linguistics, 14, 1988, 4, s. 10-34.

CLOCKSIN, W. F. - MELLISH, C. S.: Programming in Prolog. Berlin, Springer 1981. 255 s.
CSUHAJ-VARJÚ, E. - KELEMEN, J.: On the Power of Cooperation - A Regular Representation of Recursively Enumerable Languages. Theoretical Computer Science 81, 1991, s. 305-310.
DANEŠ, F.: Větné členy obligatórní, potenciální a fakultativní. In: Miscellanea linguistica, 1971, s. 131 - 138.
DANEŠ, F. - HLAVSA, Z. a kol.: Větné vzorce v češtině. Praha, Academia 1981. 270 s.
DOKULIL, M.: Tvoření slov v češtině 1. Teorie odvozování slov. Praha, nakladatelství ČSAV 1962. 264 s.
DOKULIL, M.: Tvoření slov v češtině 2. Odvozování podstatných jmen. Praha, Academia 1967. 780 s.
DOROSZEWSKI, W.: Monografje slowotwórcze. In: Studie i szkize jezykoznawcze, Warszawa, Panstwowe Wydawnictwo Naukowe 1962, s. 205-222.
DVONČ, L. a kol: Morfológia slovenského jazyka. Bratislava, Vydavateľstvo SAV 1966.
DVONČOVÁ, J.: Fyziologická fonetika. Bratislava, Slovenské pedagogické nakladateľstvo 1980. 156 s.
FEYNMAN, R.P.: To nemyslíte vážne, pán Feynman. Bratislava, Obzor 1990. 332 s.
FIKES, R. E. - NILSSON, N. J.: STRIPS: A New Approach to the Application of Theorem Proving in Problem Solving. Artificial Intelligence 2, 1971, s. 189-208.
FILIPEC, J. - ČERMÁK, F.: Česká lexikologie. Praha, Academia 1985.
FILLMORE, C. J.: The Case for Case. In: Universals in Linguistic Theory. Red. E.Bach - R. Harms. New York, Holt 1968.
FLORES, C. F.: Management and Communication in the Office of the Future. San Francisco, Hermenet Inc. 1982.
FURDÍK, J.: K syntakticko-sémantickej interpretácii slovotvorných štruktúr. Slavica slovaca, 13, 1978, s. 274-287.
Fuzzy Information, Knowledge Representation and Decision Analysis. Proceedings of the IFAC Symposium July 19-21, Marseille, 1983.
GÁL, E. - KELEMEN, J.: Myseľ, telo, stroj. Bratislava, Bradlo 1992.
GAREY, M. - JOHNSON, D.: Computers and Intractability. San Francisco, Freeman 1979.
GROF, S.: Beyond the Brain. Birth, Death & Transcendence. New York, University Press 1985.
HAJIČOVÁ, E. - KUBOŇ, P. - KUBOŇ, V.: Hierarchy of Salience and Discourse Analysis and Production. In: Proceedings of COLING '90, Helsinki, 1990.
HAJIČOVÁ, E. - KUBOŇ, V. - KUBOŇ, P.: Stock of Shared Knowledge - a Tool for Solving Pronominal Anaphora. In: Proceedings of COLING '92, Nantes, 1992, s. 127-133.
HAJIČOVÁ, E. - SGALL, P.: Automatické zpracování přirozeného jazyka. Bratislava, Elektrotechnická fakulta SVŠT 1987.

HALLIDAY, M. A. K. - HASAN, R.: Cohesion in English. London, Longman 1976.
HAVEL, I. M.: Robotika. Praha, SUTL 1980. 400 s.
HEIDEGGER, M.: What Is Called Thinking? New York, Harper & Row 1968.
HNÁTKOVÁ, M.: Inferenční procedura systemu TIBAQ II. Praha, Matematicko-fyzikálna fakulta UK 1985. HOPCROFT, J. - ULLMAN, J.: Introduction to Automata Theory, Languages and Computation. Reading, Addison-Wesley 1979. 200 s.
HORECKÝ, J.: Slovotvorná sústava slovenčiny. Bratislava, vydavateľstvo SAV 1959. 218 s.
HORECKÝ, J.: Morfematická štruktúra slovenčiny. Bratislava, vydavateľstvo SAV 1964. 194 s.
HORECKÝ, J.: Slovenská lexikológia. Bratislava, Slovenské pedagogické nakladateľstvo 1971. 249 s.
HORECKÝ, J.: Základy jazykovedy. Bratislava, Slovenské pedagogické nakladateľstvo 1977.
HORECKÝ, J.: Slovotvorná parafráza. Jazykovedný časopis, 32, 1981, s. 6-12.
HORECKÝ, J.: Ein System von Konzeptualisatoren für Künstliche Intelligenz. In: Proceedings of the Fourteenth International Congress of Linguistics, Berlin 1987, s. 1187-1189.
HORECKÝ, J.: Výskum komunikačných procesov. Jazykovedný časopis, 38, 1987, s. 47-51.
HORECKÝ, J. - BUZÁSSYOVÁ, K. - BOSÁK, J. a kol.: Dynamika slovnej zásoby súčasnej slovenčiny. Bratislava, Veda 1989. 429 s.
HUANG, X. - GUTHRIE, L.: Parsing in Parallel. New Mexico, Computing Research Laboratory of New Mexico State University 1985.
CHESTERTON, G. K.: Ortodoxie. Brno, Tom 1993. 156 s.
CHOMSKY, N.: Three models for the description of language. IRE Transactions PGIT, 2, 1956, s. 113-124.
CHOMSKY, N.: Aspects of the Theory of Syntax. Cambridge, MIT Press 1965.
KAČALA, J.: Doplnok v slovenčine. Bratislava, Vydavateľsvto SAV 1971. 277 s.
KAMP, H.: From Discourse to Logic. Acireale, Scuola estiva della linguistica computazionale 1989. 400 s.
KAPLAN, R. - BRESNAN, J.: Lexical-functional Grammar: a Formal System for Grammatical Representation. In: The Mental Representation of Grammatical Relations. Red. J.Bresnan. Cambridge, MIT Press 1982, s. 173-281.
KARTTUNEN, L.: KIMMO: A two-level morphological analyzer. Texas Linguistic Forum 22, 1983, s. 168-186.
KAY, M.: The MIND system. In: Natural Language Processing. Red. Rustin. New York, Algorithmics Press 1973, 155-188.
KAY, M. - KAPLAN, R.: Word Recognition. Palo Alto, Xerox Palo Alto Research Center 1982.
KELEMEN, J.: Multiagent symbol systems and behavior-based robots. Applied Artificial Intelligence 7 (1993), s. 419-432.

KELEMEN, J. a kol.: Základy umelej inteligencie. Bratislava, Alfa 1992. 399 s.
KITTREDGE, R. a kol.: „TAUM-73". Montreal, Université de Montreal 1973.
KUHN, T. S.: Štruktúra vedeckých revolúcií. Bratislava, Pravda, 1981.
Krátky slovník slovenského jazyka. Bratislava, Veda 1987. 587 s.
LAKOFF, G. - JOHNSON, M.: Metaphors We Live By. Chicago, University Press 1980. 242 s.
LORENZ, K.: Die angeborenen Formen möglicher Erfahrung. Tierpsychologie 5 (1943), s. 235-409.
MacLEAN, D.: To Hear the Angels Sing. Hudson, Lindisfarne Press 1980. 217 s.
MARCUS, M.: Theory of Syntactic Recognition for Natural Language. Cambridge, MIT Press 1980.
MARR, D.: Vision. San Francisco, W.H.Freeman & Co. 1982.
MATURANA, H. R.: Biology of Cognition. Dordrecht, Reidel 1980.
McCARTHY, J. - HAYES, P. J.: Some Philosophical Problems from the Standpoint of Artificial Intelligence. In: Machine Intelligence, vol. 4. Red. B. Meltzer - D. Michie. Edinburgh, University Press 1969.
MEĽČUK, I. A.: Opyt teorii lingvitičeskich modelej. Moskva, Nauka 1974. 314 s.
MIKO, F.: Rod, číslo a pád podstatných mien. Bratislava, Vydavateľstvo SAV 1962. 254 s.
MINSKY, M. - PAPERT, S.: Perceptrons. Cambridge, MIT Press 1969. 40 s.
MINSKY, M.: A framework for representing knowledge. In: The Psychology of Computer Vision. Red. P. H. Winston. New York, McGraw-Hill 1975.
MINSKY, M.: The Society of Mind. New York, Simon & Schuster 1986. 332 s.
MISTRÍK, J.: Štylistika slovenského jazyka. Bratislava, Slovenské pedagogické nakladateľstvo 1977. 444 s.
MLACEK, J.: Slovenská frazeológia. Bratislava, Slovenské pedagogické nakladateľstvo 1977. 116 s.
Mluvnice češtiny 1. Fonetika, fonologie, morfonologie a morfematika, tvoření slov. Praha, Academia 1986. 557 s.
Mluvnice češtiny 3. Skladba. Praha, Academia 1987. 738 s.
NALIMOV, V. V.: Verojatnostnaja model' jazyka. Moskva, Nauka 1979.
NALIMOV, V. V.: Realms of the Unconscious: The Enchanted Frontier. Philadelphia, ISI Press 1982. 320 s.
NEWELL, A.: Physical Symbol Systems. Cognitive Science 4, 1980, s. 135-183.
NEWELL, A. - SIMON, H. A.: Computer science as empirical inquiry:
symbols and search. Communications of the ACM 19, 1976, s. 113-126.
NIDA, E. A.: Computational Analysis of Meaning. Hague, Mouton Publishers, 1975. 272 s.
NILSSON, N. J.: Principles of Artificial Intelligence. Berlin, Springer Verlag 1982. 400 s.
NOVACKÝ, M. - CZAKO, M.: Základy etológie. Bratislava, Slovenské pedagogické nakladateľstvo 1987. 250 s.
OLIVA, K.: *Stručný přehled syntaktické analýzy češtiny. MFF UK Praha.*

ONDREJOVIČ, S.: Medzi slovesom a vetou. Problémy slovesnej konverzie. Bratislava, Veda 1989. 123 s.
ORAVEC, J.: Väzba slovies v slovenčine. Bratislava, Vydavateľstvo SAV 1967. 388 s.
ORAVEC, J. - BAJZÍKOVÁ, E. - FURDÍK, J.: Súčasný slovenský spisovný jazyk. Morfológia. Bratislava, Slovenské pedagogické nakladateľstvo 1984. 227 s.
ORAVEC, J. - BAJZÍKOVÁ, E.: Súčasný slovenský spisovný jazyk. Syntax. Bratislava, Slovenské pedagogické nakladateľstvo 1982. 261 s.
OSGOOD, A. F. - SANN, G. J. - TANNENBAUM, P. H.: The Measurement of Meaning. Urbana, 1957.
PÁLEŠ, E.: ANDULKA. Papagájovací stroj angličtiny. Dokumentácia k programu. Bratislava, MFF UK 1987. 12 s.
PÁLEŠ, E.: SAMANTA. Sémantická analýza. Práca ŠVOČ. Bratislava, MFF UK 1988. 77 s.
PÁLEŠ, E.: SAPFO - systém pre komunikáciu v prirodzenom jazyku I. Morfológia, lexikálna sémantika, SRL. Diplomová práca. Bratislava, Univerzita Komenského 1988. 207 s.
PÁLEŠ, E.: SAPFO - systém pre komunikáciu v prirodzenom jazyku II. Algoritmy syntakticko-sémantickej analýzy. Diplomová práca. Bratislava, Univerzita Komenského 1989. 70 s.
PÁLEŠ, E.: Spolupráca syntaxe a sémantiky pri analýze flektívnych jazykov. In: SOFSEM '89. Ždiar, 1989, s. 53-56.
PÁLEŠ, E.: Sémantické roly slovenských slovies. Jazykovedný časopis 41, 1990, s. 30-48.
PÁLEŠ, E.: Co-operation of Syntax and Semantics in Flexive Languages. Journal of Experimental & Theoretical Artificial Intelligence, 2, 1990, s. 1-24.
PÁLEŠ, E.: A Computational Model of Language as a Whole. In: Proceedings of the 7th International Meeting of Young Computer Scientists. Magdeburg, University Press 1992, s. 172-179.
PÁLEŠ, E.: Počítačový model jazyka ako celku. In: Kolokvium mladých jazykovedcov I. Bratislava, SJS pri SAV 1992, s. 15-20.
PÁLEŠ, E.: Natural Language Paraphraser - A Computer Tool for Linguists. Computers & Artificial Intelligence 12, 1993, 5, s. 491-510.
PÁLEŠ, E.: Slovotvorná parafráza na počítači. Celostné modelovanie jazyka. Jazykovedný časopis 44, 1993, 2, s. 108-125.
PÁLEŠ, E.: Contents, Meaning & Form. A Hierarchic Paraphrasing Engine. In: Proceedings of th 6-th International conference on Artificial intelligence and information-control systems of robots. Red. I. Plander. V tlači, 1994.
PALMER, M. S.: Semantic Processing for Finite Domain. Cambridge, University Press 1990. 199 s.
PANEVOVÁ, J.: Formy a funkce ve stavbě české věty. Praha, Academia 1980. 224 s.
PANEVOVÁ, J.: Kritéria pro zjišťování významu vět. Slovo a slovesnost, 1978.
PATAŇJALI: Die Yoga-sutras. In: Vivekananda, 1988.

PEREIRA, F. - WARREN, D. H. D.: Definite Clause Grammars for Language Analysis - a survey of the formalism and a comparison with augmented transition networks. Artificial Intelligence, 13, 1980, s. 231-278.

PISARČÍKOVÁ, M. - MICHALUS, Š.: Malý synonymický slovník. Bratislava, Slovenské pedagogické nakladateľstvo 1988. 252 s.

POINCARÉ, H.: Letzte Gedanken. Leipzig, Akademische Verlagsgesellschaft 1913. 261 s.

POPPER, M. - GYÁRFÁŠ, F.: CODEX - A Computer-based Diagnostic Expert System. In: Artificial Intelligence and Information-Control Systems of Robots. Red. I. Plander. Amsterdam, North-Holland 1984, s. 297-300.

POPPER, M. - KELEMEN, J.: Expertné systémy. Bratislava, Alfa 1989.

PUSTEJOVSKY, J: The Generative Lexicon. Computational Linguistics, 17, 1991, 4.

RADHAKRISHNAN, S.: Die Bhagavadgita. Baden-Baden, Holle 1958. 448 s.

RAJNEESH, O.: Psychology of the Esoteric. Cologne, Rebel Publishing House 1982. 178 s.

REVZINA, O. G.: Struktura slovobrazovateľnych polej v slavianskich jazykach. Moskva, Izdateľstvo Moskovskogo universiteta 1969. 156 s.

RISTAD, E. S. - BERWICK, R. C.: Computational Consequence of Agreement and Ambiguity in Natural Language. Journal of Mathematical Psychology 33, 1989, s. 379-396.

RITCHIE, G. D.: Computational Grammar - An AI Approach to Linguistic Description. Brighton, Harvester Press 1980. 330 s.

ROUNDS, W. C.: LFP: A Logic for Linguistic Descriptions and an Analysis of its Complexity. Computational Linguistics, 14, 1988, 4, s. 1-9.

SALTON, G.: Automatic Information Organization & Retrieval. New York, Mac Graw-Hill 1968.

SALTON, G.: Prieskumný informačný systém. Bratislava, Alfa 1981. 587s.

SALTON, G. - McGILL, M.: An Introduction to Modern Information Retrieval. New York, McGraw-Hill 1983.

SEGOND, F. - JENSEN, K.: Un Systeme Syntaxique et Semantique Integre pour la Comprehension du Langage Naturel. In: COLING 1992, s. 890-895.

SEJÁKOVÁ, J.: Slovotvorná synonymia v dvojici typu žltý - zožltnutý. Jazykovedný časopis, 39, 1988, s. 51-62.

SGALL, P.: Contributions to Functional Syntax, Semantics and Language Comprehension. Praha, Academia 1984. 379s.

SGALL, P.: Lingvistika na rozcestí? Jazykovedný časopis, 29, 1988, s. 144-152.

SGALL, P. - HAJIČOVÁ, E. - PANEVOVÁ, J.: The Meaning of the Sentence. Praha, Academia, 1986.

SHASTRI, L.: Evidential Reasoning in Semantic Networks: A Formal Theory and its Parallel Implementation. Rochester, Department of Computer Science, University of Rochester 1985.

SHELDRAKE, R.: The Presence of the Past. London, Fontana 1989. 391 s.

SIMON, H. A.: Search and reasoning in problem solving. Artificial Intelligence 21, 1983, s. 7-29.
SMIEŠKOVÁ, E.: Malý frazeologický slovník. Bratislava, Slovenské pedagogické nakladateľstvo 1989. 270 s.
VELARDI, P. - PAZIENZA, M. T. - FASOLO, M.: How to Encode Semantic Knowledge: A Method for Meaning Representation and Computer-Aided Acquisition. Computational Linguistics, 17, 1991, s. 153-170.
SMADJA, F.: Retrieving Collocations from Text: Xtract. Computational Linguistics 19, 1993, s. 143-177.
SMETÁČEK, V.: Systém sémantického analyzátoru a báze BALEX. Pracovný materiál ÚVTEI. Praha, ÚVTEI 1988. 25 s.
ŠEFRÁNEK, J.: Fillmorova pádová gramatika. Pracovný materiál ÚVTVŠ. Bratislava, ÚVTVŠ 1970. 12 s.
TESNIÉRE, L.: Eléments de syntaxe structurale. Paris, 1959.
TIBENSKÁ, E.: Návrh na špecifikáciu subjektového participanta v SSJ. Pracovný materiál JÚĽŠ SAV. Bratislava, JÚĽŠ 1988.
TIBENSKÁ, E.: Kauzatívne konštrukcie z hľadiska sémantickej štruktúry. Jazykovedný časopis, 40, 1990, s. 21-32.
TURING, A. M.: Computing machinery and intelligence. Mind 59, 1950, s. 433-460.
VAPNIK, V. M. - ČERVONENKIS, A. Ja.: Teorija raspoznavanija obrazov. Moskva, Nauka 1974.
VIJAYASHANKAR, K. - JOSHI, A. K.: Some Computational Properties of Tree Adjoining Languages. In: Proceedings of the 23rd Meeting of Association of Computational Linguistics, Chicago, 1985, s. 82-89.
VINOGRADOV, V. V.: Ob osnovnych tipach frazeologičeskich jedinic v russkom jazyke. In: A.A.Šachmatov. Moskva-Leningrad, 1947, s. 339-364.
VIVEKANANDA, S.: Raja-Yoga. Freiburg, Bauer 1988, 286 s.
WAHLSTER, W.: Discourse and User Models. Pracovný materiál. Praha, Summer school in computational linguistics 1991.
WEINREICH, U.: Travels in Semantic Space. Word, 14, 1958, s. 346-366.
WEIZENBAUM, J.: ELIZA. Communications of the ACM 9, 1966, s.36-45.
WHORF, B. L.: The Relation of Habitual Thought and Behavior to Language. In: Language, Thought and Reality. Red. B. L. Whorf. New York, Wiley 1956.
WILENSKY, R. - CHIN, D. N. - LURIA, M. - MARTIN, J. - MAYFIELD, J. - WU, D.: The Berkeley UMX Consultant Project. Computational Linguistics 14, 1988, 4, s. 1-84.
WINOGRAD, T.: Understanding Natural Language. New York, Academic Press 1972.
WINOGRAD, T.: Five Lectures on Artificial Intelligence. In Linguistic Structures Processing. Red. A. Zampolli. Amsterdam, North Holland 1977.
WINOGRAD, T.: Language as a Cognitive Process, vol. I - Syntax. Reading, Addison-Wesley 1983.
WINOGRAD, T. - FLORES, F.: Understanding Computers & Cognition. Reading, Addison-Wesley 1987.

WINSTON, P. H.: Learning Structural Descriptions from Examples. In: The Psychology of Computer Vision. Red. P. H. Winston. New York, Mac Graw-Hill 1975.

WINSTON, P. H.: Artificial Intelligence. Reading, Addison-Wesley 1977. 439 s.

WINSTON, P. H.: Learning by Creating and Justifying Transfer Frames. Pracovný materiál. Cambridge, MIT AIL 1977.

WOODS, W.: LUNAR - Lunar Rocks in Natural Language. Explorations in natural language question answering. In: Linguistic Structures Processing. Red. A. Zampolli. Amsterdam, North Holland 1977.

ZHELEZNYAKOV, M. M. - NEVLEVA, T. N. - NOVITSKAJA, I. M. - SMIRNOVA, L. N. - TSEYTIN, G. S.: An Attempt towards „Text -> Reality" Models Employing Association Nets. Pracovný materiál. Leningrad 1986.

ZNÁM, Š.: Teória grafov. Bratislava, MFF UK 1980. 100 s.

Konferencie

[1] First International Dialogue: Transition to Global Society. Landegg Academy, Landegg, Schweiz, 3-9 júla 1990.

[2] The 12-th International Transpersonal Conference: Science, Spirituality & the Global Crisis: Toward a World with a Future. Praha, 20-25 júna 1992. .

[3] COLING. The fifteenth International Conference on Computational Linguistics, Nantes, júl 1992.

Natural Language Paraphraser - A Computational Tool for Linguists

SUMMARY

SAPFO (Sense Apprehending Frame Operator) is a natural language paraphrasing system. The user types in a sentence or a phrase in Slovak and the program replays with a set of Slovak sentences (phrases) of different forms, but of a meaning close to the original one. This system serves as an experimental (and simplified) model of human linguistic consciousness and has been based on a psycholinguistic experiment with listening and re-narrating of short stories.

A lot of approaches have been proposed in computational linguistics. Not many of them have been implemented succesfully. One of the reasons for this is, that particular solutions for various specialized problems are proposed „out of context", i.e. without *having in mind the whole of the system*. Then it turns out, that the particular procedures are either not compatible among themselves, or that they shift some fundamental problem over to another parts of the system, or they expect a kind of input, which is not available.

On the other hand, there are some succesfully applied systems, but they do not meet the demand of wholeness either. They usually work only thanks to a radical restriction of discourse domain, or partially thanks to a sophisticated collection of tricks, rather than being a valid model of language consciousness.

SAPFO is designed to be a model of language as a whole. It includes all subsystems of language - morphology, syntax, semantics, word-formation, phraseology, stylistics - and enables to watch them at work in mutual interaction. The program consists of several parallely working a communicating independent modules (agents). Each one of them represents one of the subsystems of language.

The effect of *synergy* arises in such a holistic (emergent or multi-agent) system and helps to reduce the computational complexity and to enhance the cognitive power of the whole system.

The morphological module contains a complete declarative description of Slovak morphology (declination, conjugation, comparison etc.). It is able to analyse or to generate any Slovak word-form together with its grammatical categories.

The module of syntax is an ATN with a strong semantic component. It covers a broad scale of Slovak syntactic phenomena and works in close communication with semantic collocation functions.

The semantic module consists of lexical-semantic information, valency frames of verbs, nouns, adjectives and others. It is competent to decide the semantic admissability of any syntagm and communicates with other modules in the form of five collocation functions VAL, LVAL, RVAL, CVAL, DVAL (verb-valency, left and right valency of nouns, inner valency of composites and derivatives).

The module of derivatology contains a powerful description of Slovak composition and derivation. Several hundred dynamic rules are covering the major part of

Slovak word-formation. This module enables to unfold compound words into sentences or NP's and vice versa.

The module of phraseology enables to exchange certain parts of the sentence with equivalent phraseological expressions and vice versa.

The module of stylistics controls the stylistic consistency of the sentence. It enables to form or transform a sentence according to different styles or dialects.

These modules in cooperation are able to perform in total about 70 different types of manipulations with form and meaning. Among them there are transformations (transpositions of grammatical cases and other surface forms, like passivization, nominalization, composition, derivation, participalization, infinitivization, apposition, subordination, coordination etc.), domination (vertical rearrangement of the sentence), transfigurations (perspectivization of semantic figures), transmutations (addition or deletion of elementary components of meaning, like synonymization, antonymization, magnonymization, deminutivization etc.). The system enables to perform also changes of intonation, topic-focus articulation, stylization and some poetic figures.

The paraphrasing means, which are to be applied and watched, can be choosen by the user from a menu-table of language manipulations and switched on and off one by one or in groups.

The system serves as a computational tool for linguists, who can use the program for testing and verification of various linguistic hypotheses on large corpora. Experience with SAPFO can contribute in the future also to the development of applied information retrieval systems in natural language.

This system is unique in Slavonic languages.

Der Paraphraser von Slowakisch - Ein Computer-werkzeug für linguistische Modellierung

ZUSAMMENFASSUNG

SAPFO (Sense Apprehending Frame Operator) ist ein System für Paraphrasieren der natürlichen Sprache. Der Benützer tippt einen Satz oder eine Phrase in Slowakisch ein, und das System antwortet mit Generierung von Slowakischen Sätzen und Phrasen, die zwar verschiedene Formen annehmen können, aber in der Bedeutung dem ursprünglichem Satz nahe stehen. Als Grundlage für das System diente ein Psycholinguistisches Experiment mit Zuhören und Wiedererzählen von kurzen Erzählungen. Es soll eine (vereinfachte) Art Modell des menschlichen Sprachbewußtseins darstellen.

Eine Menge von Ideen ist vorgeschlagen worden in den veragangenen Jahren in Computerlinguistik. Nur wenige von ihnen vermochte man erfolgreich implementieren. Einer der Gründe dafür ist, daß die partikuläre Lösungen für verschiedene Spezialprobleme sind „ohne kontext" vorgeschlagen worden, d.h. ohne *das Ganze des Systems im Sinn habend.* Dann stellt sich heraus, daß die Teilprozeduren untereinander nicht kompatibel sind, oder daß sie die Verantwortung für irgendwelches wichtiges Problem auf andere Teile des Systems außerhalb sich selbst verschieben, oder daß sie eine Art Eingang erwarten, das nicht vorhanden ist oder gar sein kann.

Dann gibt es auch erfolgreiche Applikazionen, Systemen, die aber aus einem anderen Grund die Bedingung der Ganzheitlichkeit nicht erfüllen. Diese arbeiten überwiegend dank einer radikalen Verengung der Diskursdomäne, und teilweise vielleicht auch dank einer klugen Kollekzion von Tricks, und können kaum als allgemeine Modelle des Sprachbewußtseins dienen.

SAPFO ist als ein Modell der gesamten Sprache als eines Ganzen gebaut. Es schließt alle Subsystemen der Sprache - Morphologie, Syntax, Wortbildung, Semantik, Phraseologie, Stylistik u.a. - ein, und ermöglicht diese in der gegenseitigen Wechselwirkung an der Arbeit zu beobachten. Das Programm besteht aus mehreren paralell arbeitenden und untereinander kommunizierenden Modulen (Agenten); jeder von denen je ein Subsystem der Sprache darstellt.

In so einem ganzheitlichen Modell entsteht ein synergisches Effekt, der die Rechenkomplexität zu vermindern hilft und die Gesamtleistung des Systems steigert.

Das morphologische Modul enhält eine vollständige Beschreibung der Slowakischen Formenlehre (Deklinazion, Konjugation, Steigerung usw.). Es ist im Stande beliebige Slowakische Wortform mit ihren grammatischen Kategorien zu bilden oder zu analysieren.

Das syntaktische Modul ist ein Übergangsnetz (ATN) mit einer starken semantischen Komponente. Es bedeckt eine breite Skala von Slowakischen syntaktischen Phenomäne und arbeitet eng mit den semantischen Kollokationsfunkzionen zusammen.

Das semantische Modul besteht aus der lexikal-semantischen Information, Valenzrahmen von Verben, Substantiven u.a. Es entscheidet über die semantische Zulässigkeit der Syntagmen und kommuniziert mit anderen Modulen durch die fünf Kollokationsfunkzionen VAL, LVAL, RVAL, CVAL, DVAL (Verbvalenz, linke und rechte Nominalvalenz, Valenz innerhalb von Kompositen und Derivaten).

Das Wortbildungmodul enthält umfangreiche Beschreibung der Slowakischen Komposition und Derivation. Mehrere hundert von Regeln bedecken den größten Teil der Slowakischen Wortbildung. Dieses Modul ermöglicht die Kompositen und Derivaten wieder in vollständige Sätze (oder Phrasen) und umgekehrt zu transformieren.

Das Modul der Phraseologie ermöglicht einige Satzteile mit einem phraseologischen Äquivalent und umgekehrt zu ersetzen.

Das Modul der Stylistik überwacht die stylistische Konsistenz der Sätze. Es ermöglicht die Sätze in Eintracht mit verschiedenen Stylen oder Dialekten (um)zu formen.

Diese Modulen zusammenarbeitend sind im Stande etwa 70 verschiedene Arten von Sprachmanipulationen durchzuführen. Unter diese gehören z. B. Transformationen (Transpositionen von grammatischen Fällen und anderen Oberflächenformen, wie Passivization, Nominalization, Komposition, Derivation, Partizipalization, Infinitivization, Apposition, Subordination, Koordination usw.), Domination (vertikale Umformung des Satzes), Transfiguration (Perspektivization der semantischen Figuren), Transmutation (Zufügung oder Auslassung einer elementaren Komponente der Bedeutung, wie Synonymization, Antonymization, Magnonymization, Deminutivization usw.). Außerdem ermöglicht das System die Intonation, Wortfolge und den Styl zu verändern.

Der Benützer kann die Paraphrasierungsmittel, die er applizieren und beobachten will, in einer Menu-tafel von Sprachmanipulazionen einzeln oder in Gruppen ein- und ausschalten.

SAPFO dient als ein Computerwerkzeug für die Sprachwissenschaftler, die bei Hilfe dieses Systems verschiedene linguistische Hypothesen auf einem großen Korpus testen und verifizieren können. Die Erfahrung mit dem SAPFO-experiment könnte später auch zur Bau von applizierten Informationssystemen in natürlichen Sprache beitragen.

Ein solches System ist bisher einzigartig in Slawischen Sprachen.

SAPFO
A HIERARCHIC PARAPHRASING ENGINE

Abstract

SAPFO is a computer-paraphraser of Slovak. This paper gives an outline of the hierarchic organization of various paraphrasing means of natural language. From a deep representation, the paraphraser is able to generate hierarchic families of paraphrases, expressing the same contents in different forms. The deep contents undergoes seven processes until it becomes a surface form: *selection, mutation, figuration, domination, formation, focation and intonation*. Different choice of parameters during these seven stages gives different resulting surface phrases. The hierarchic systematization of paraphrases resulted from a psycholinguistic experiment with re-narration of stories by human speakers and was then implemented on an IBM PC.

SAPFO

SAPFO (Sense Apprehending Frame Operator) is a computational model of Slovak language as a whole. It includes morphology, syntax, semantics, word-formation, phraseology, stylistics and lexical paradigmatics. The system enables to watch and to study the behavior of all subsystems of natural language at work in their mutual interaction.

SAPFO is implemented in ARITY-PROLOG and runs on an IBM PC AT 386 DX computer. It consists of nine communicating and co-operating modules, total of 200 K code plus 400 K lexical information.

A complete description of representation and interpretation formalisms of SAPFO can be found in (Páleš, 1993b). The co-operation of the syntactic and semantic module is discussed in more detail in (Páleš, 1990b) and the interaction between syntax and word-formation in (Páleš, 1993a).

SRL. A box-of-bricks representation

SRL (Sentence Representation Language) is a symbolic language for representation of contents and meaning of natural language phrases. The meaning of each Slovak phrase is represented by a semantic net (or an equivalent frame), which is built up from a combination of SRL-symbols. The inventory of SRL-symbols contains three finite sets of symbols, i.e. three types of 'bricks':

1. A set of word-roots, which are derivationally irreducible, i.e. they constitute the root of the whole word-formation nest. For example *gold-* and *love-* are the

innermost roots of the word-formation nests *golden, goldish, goldbearing, goldsmith...* and *lover, loveless, lovely, loveliness, loveable, love-song...*[1]

2. A set of 88 semantic cases, like AGENT, PATIENT, BENEFACTOR, INSTRUMENT, ADDRESSEE, ORIGIN, LOCATION, DESTINATION, ATTRIBUTE, CAUSE, POSSESSOR etc.

3. A set of 66 'components of meaning'. A piece of meaning becomes a component of meaning, if it functions regularly in a certain language. Among these 66 components of meaning there are only the most important ones, which have also a specific formal means of expression in a particular language and are needed to generate paraphrases correctly.

Examples: *to pour out = to pour +ex, I will have done = I am doing +fut +perf, forgeable = forged +abil, colorless = colored +neg, foliage = leaf +pl, lady = lord +fem, reddish = red +min, lass = girl +poet.*

Here *+ex, +fut, +perf, +abil, +neg, +pl, +fem, +min, +poet* are components of meaning.

The lexicon of SAPFO contains 300 complete word-formation nests now and may easily be expanded. A complete list of semantic cases and their meanings has been published in (Páleš, 1990a). The set of components of meaning with a commentary can be found in (Páleš, 1993a).

The semantic representation of a phrase is to be constructed as follows:

1. Word-roots constitute the set of node labels.
2. Semantic cases constitute set of arc labels.
3. The components of meaning may be adjoined to the nodes.

For example the sentence

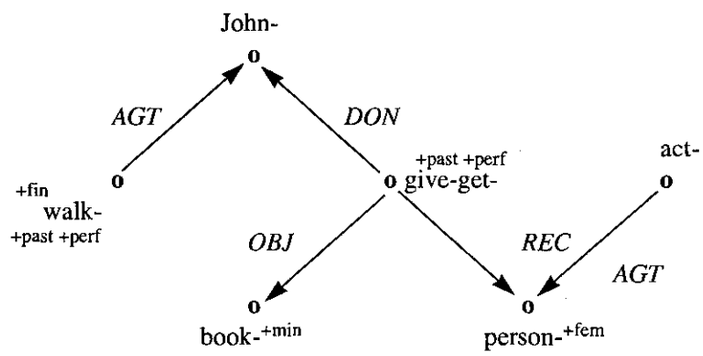
John stopped walking and gave the booklet to the actress. (1)

is represented by a frame:

```
John-
  «AGENT walk- +fin +past
  «DONOR give-get- +past +perf                    (2)
            OBJECT» book- +min
            RECIPIENT» person- +fem
                        «AGENT act-
```

[1] Compound words may contain up to two or three roots - i.e. nouns, verbs or adjectives with full meaning - together with several affixes with a partial meaning only.

or by an equivalent semantic net:



(3)

Collocation functions

The core of the system is a large ATN and a unification grammar, containing formal patterns of semantic syntax and word-formation. The process of parsing is controlled by five *collocation functions*:

VAL (verb valency)
LVAL (left valency of nouns)
RVAL (right valency of nouns)
CVAL (inner compositionality of composites)
DVAL (inner compositionality of derivatives)

A collocation function defines the compositionality of lexical units and at the same time provides an algorithm for transition between form and meaning and vice versa. Examples (simplified):

VAL(motion-verb,0,nominative,animate) = AGENS

Jano^{AGT} cestuje John^{AGT} travels

VAL(motion-verb,do,genitive,location) = DESTINATION

cestuje do Prahy^{DES} he left for Prague^{DES}

VAL(motion-verb,0,instrumental,instrument) = INSTRUMENT

cestuje autom^{INS} he travels by car^{INS}

LVAL(rel-adj,-ický,location,noun,_,movable) = ORIGINATIVE

indické^{ORI} korenie indian^{ORI} pepper

LVAL(rel-adj,-ný,location,noun,_,unmovable) = LOCATIVE

južná^{LOK} veža southern^{LOC} tower

RVAL(noun,concrete,z,genitive,noun,materialium) = MATERIAL

prsteň zo zlata^{MAT} ring of gold^{MAT}

RVAL(noun,event,z,genitive,noun,abstract) = CAUSATIVE

dar z lásky^{KAZ} gift of love^{CAS}

CVAL(noun,instr,verb,-ec,person) = INSTRUMENT, AGENS

delo^{INS}strelec^{AGT} canon^{INS}shooter^{AGT}

DVAL(noun,mater,verb,_,location) = OBJECT, LOCATIVE

pivo^{OBJ}var ^{LOK} beer^{OBJ}cook ^{LOC} (brewery)

Using collocation functions bothways together with unification grammar enables to analyse and generate natural language. Collocation functions automatically resolve many cases of polysemy and structural ambiguity.

Cooperation of syntax and semantics in a flective language. An example

Let us illustrate the whole parsing process on an example, a typical one for Slovak. We will watch the co-operation of the syntactic and semantic modules of SAPFO on the background of lexical-semantic knowledge.

Parsing the more transparent structure of English, backtrackings, ambiguities dragging the parser into dead ends are typical. The Slovak parser, though apparently more complicated, moves on more deterministically, showing its power when investigating the involved system of cases and various morphological and semantical marks elegantly decides the semantic roles or chooses between two meanings of a polysemical word.

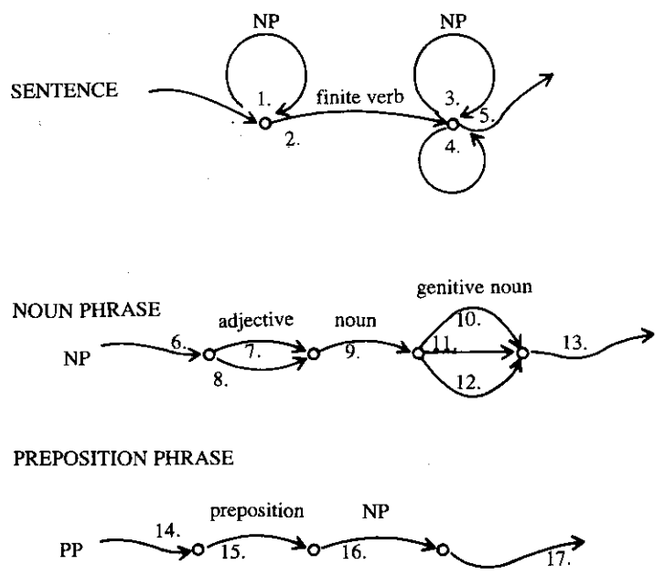


Figure 1. A simplified ATN.

Table 1. Conditions and actions for the ATN in figure 1.

arcs 1, 2, 3	Agreement of subject (the only NP in nominative) with the main verb in person, number and animateness.
arcs 7, 9	Adjective-noun agreement in case, gender, number and animateness.
arc 5	The head of the automat at the end of the input string.
arcs 1, 3, 4	The VAL function tries to assign a semantic role to the NP or PP participant, in the valency field of the main verb.
arc 7	The LVAL function assigns the role to the left-standing attribute of the head noun of an NP.
arcs 10, 12	The RVAL function assigns a semantic role to the right-standing attribute of an NP in the valency of the head noun on arc 9.

Suppose we want to parse the sentence:

OBJECT	MAIN VERB	SUBJECT	PP
Južné krídlo	*hradu postavili*	*Florentskí majstri*	*z guľatých skál.*

southern wing (of the) castle (has been) built (by) Florentine masters of globular stones.

The object - as the information already known - is at the beginning of the sentence. The subject and the PP are standing at the end, being the new or expected information. We will use the simple ATN in figure 1. The conditions and actions to this ATN are given in table 1. The morphological analysis gives the following output:

južné	=	relation-adjective *južný* (southern), gender female or neutral, case nominative or accusative, number sg. or pl., animate or inanimate, derived from the noun *juh* (south), by the suffix *-ný*.
krídlo	=	noun *krídlo* (wing), gender neutral, case nominative or accusative, number sg., inanimate, unmovable (the wing of a castle) or movable (the wing of a bird).
hradu	=	noun *hrad* (castle), gender male, case genitive or dative, number sg., inanimate, unmovable
postavili	=	l-participle of the verb *postaviť* (to build), person 3., number pl., past tense, positive, valency frame no. 20.
Florentskí	=	relation-adjective *florentský* (florentine), gender male, case nominative, number pl., animate, derived from the noun *Florencia* (Florence), location, by the suffix *-ský*.
majstri	=	noun *majster* (master), gender male, case nominative, number pl., animate, non-derived.
z	=	genitive preposition *z* (from)
guľatých	=	adjective *guľatý* (globular), gender male, female or neutral, case genitive or local, number pl., animate or inanimate, desubstantive from *guľa* (globe), derived by the suffix *-atý*.
skál	=	noun *skala* (rock), gender female, case genitive, number pl., inanimate, non-derived.

Note that the morphological output is ambiguous, but each word remains within its word class. No word is a verb and a noun at the same time. The leftmost non-modal verb - *postavili* - will be the main verb. That is, all the participants in the sentence are standing in the valency of the following frame (the first fourteen lines):

VAL(postaviť,0,nominative,animate) = AGENT
VAL(postaviť,0,nominative,instrument) = INSTRUMENT
VAL(postaviť,0,dative,animate) = BENEFACTOR
VAL(postaviť,0,accusative, concrete) = OBJECT
VAL(postaviť,0 instrumental,instrument) = INSTRUMENT
VAL(postavi5,0,instrumental,abstract) = MODUS
VAL(postaviť,z,genitive,material) = MATERIAL
VAL(postaviť,do,genitive,material) = MATERIAL
VAL(postaviť,na,local,instrument) = INSTRUMENT
VAL(postaviť,na,accusative,_) = MOTIVANT
VAL(postaviť,pre,accusative,animate) = BENEFACTOR
VAL(postaviť,za,accusative,animate) = SUBSTITUAL

VAL(postaviť,za,instrumental,location) = LOCATIVE
VAL(postaviť,za,genitive,nature force) = CONDITIONAL

The NP-interpretation rules, we will need, are:

1) LVAL(rel-adj,loc-desubst,-ný,unmov,_,_) = LOCATIVE
2) LVAL(rel-adj,loc-desubst,-ný,mov,_,_) = ORIGINATIVE
3) RVAL(mov,_,_,0,genitive,mov,_,_) = PARTITIVE
4) RVAL(unmov,_,_,0,genitive,unmov,_,_) = PARTITIVE
5) LVAL(adj,loc-desubst,-ský,mov,_,_) = ORIGINATIVE
6) RVAL(inanimate-concrete,_,_,z,genitive,material,_,_) = MATERIAL
7) LVAL(rel-adj,desubst,-atý,concrete,_,_) = RELATOR

We can start parsing. We know the main verb already - so we can begin to build the semantic tree:

MANUAL_CONSTRUCTIVE_PROCESS: *postaviť* (to build)

The sentence does not begin with a verb (arc 2). We must parse an NP (arc 1). By means of a precomputation we find the head of this NP first - *krídlo*. The grammatical categories of *južný* (on arc 7) and the head unify into: gender neutral, number sg., inanimate, case nominative or accusative. The semantic role of *južný* depends - according to the NP rules 1 and 2 - on the movability of *krídlo*. But *krídlo* is a polysemical noun. The parser does not know, whether *south* is the LOCATION of the castle's wing or its ORIGIN (as in *southern fruit*). It must consider both possibilities.

We could leave the NP through the arc 11 and try to identify *hradu* as dative object (as a next NP on arc 1). But the third line of the valency frame prevents this, failing on the demand of animateness of the BENEFACTOR. Therefore, *hradu* is in the genitive case, being the right-standing attribute of *krídlo* (arc 10).

In the rules 3 and 4 the RVAL function embodies the knowledge, that two objects, the one being a part of another, are either both movable or both unmovable. Thus the unmovability of *hrad* (castle) backwardly implies the unmovability of *krídlo* (wing) and this decides the role of *južný* (southern) as LOCATIVE. We have the NP parsed:

HEAD: *krídlo* (wing)
LOCATIVE: *juh* (south)
PARTITIVE: *hrad* (castle)

The ambiguity still remains whether this NP is in nominative (being a subject) or in accusative (object). A pure syntactic parser would recognize later that it cannot be the subject, because it does not agree with the main verb in number. But the first two lines of the currently used valency frame decide this immediately. The

castle's wing is an unmovable location. From these characteristics, neither 'animateness' nor 'instrument'features are inferrable, which are the conditions for subject roles of AGENT or INSTRUMENT.

Further, we traverse the main verb *postavili* (arc 2) saving its grammatical categories. They will later agree with the person and number of the subject *majstri*. Now, a PP does not follow (arc 4). We expect and find the NP *Florentskí majstri* (arc 3). The suffix *-ský* indicates (NP-rule 5) that the location *Florencia* stands for the ORIGINE of the masters (movable persons), not their LOCATION. The adjective-noun agreement works, the NP is in nominative, animate and obtains the semantic role AGENT (according to the first line of the valency frame):

AGENT: *majstri* (masters)
ORIGINATIVE: *Florencia* (Florence)

The PP *z guľatých skál* follows. Using the rule 7 we get

HEAD: *skaly* (stones)
RELATOR: *guľa* (globe)

The suffix *-atý* has a specific meaning - „to be like something" (globular = like a globe). This PP could have extended the previous NP. But the RVAL function (rule 6) does not allow this. The 'Florentine masters' cannot be 'of globular stones'. The seventh line of the valency frame enables to join this PP as subordinated to the main verb, with the semantic role of MATERIAL. The final analysis looks as follows:

MANUAL_CONSTRUCTIVE_PROCESS: *postaviť* (to build)
AGENT: *majstri*
ORIGINATIVE: *Florencia* (Florence)
PATIENT: *krídlo* (wing)
OBJECT: *krídlo* (wing)
LOCATIVE: *juh* (south)
PARTITIVE: *hrad* (castle)
MATERIAL: *skaly* (stones)
RELATOR: *guľa (globe)*

A computer-paraphraser of Slovak sentences

SAPFO can work as a paraphraser of Slovak sentences and phrases: The user types in a Slovak sentence, phrase or a single word. The system replies with a list of paraphrases, i.e. sentences and phrases which express the same contents through a different form.

The task of the system is to give all grammatically correct and meaningful paraphrases; nothing more and nothing less than this. Unfortunately, the system usually generates some semantically unplausible paraphrases, and on the other hand fails to generate some desirable ones. Our goal is to minimize the error, i.e. the number of wrongly generated plus correct but non-generated paraphrases.

The system first generates the mother deep structure: it splits the meaning of the original sentence into elementary SRL-symbols and components of meaning. From this deep representation it generates all the sister surface forms.

For instance, the sentence (1) is first translated into the deep structure (3) and then a list of 'similar' sentences is generated. All of them are paraphrases, i.e. different surface forms of expression for the same event (3):

John stopped walking and gave the booklet to the actress.
The walker stopped and gave the small book to the actor.
Having stopped walking, John gave the book to the woman, who acts.
When giving the book away, John stopped his walk.
John was giving the booklet to a woman. He was not walking.
The actress has got the book from John, who stopped walking.
The actress was given the booklet by John.
The book has been John's gift for the artist.
The book, which was given to the actress by John, was not big.
John's book was small...

The ability to paraphrase sentences as a criterion of 'understanding'

If a teacher wants to find out whether a pupil has understood the matter of instruction, he asks the pupil to explain it in his own words. If he fails to do that, it is obvious, that the pupil has been only memorizing what he has heard, but he does not understand the contents.

In the same way, the (un)ability of a computer to paraphrase sentences correctly makes the level and extent of its „understanding of language" evident.

A psycholinguistic experiment

Along with the development of the paraphraser, a psycholinguistic experiment with paraphrasing of stories by human speakers has been carried out (see Páleš, 1993b). We have gathered a lot of material about the variety of expression forms attributed to the same contents by humans.

Generally, the set of paraphrases of one phrase is very huge. It may contain hundreds, thousands or more paraphrases. Two paraphrases may differ in different extent and in different direction. Some differ only in tonal stress, focus or word-order, some arise by transposition of syntactic elements and grammatical cases or by different choice of lexis. There are paraphrases, which change the meaning or

affect the very contents of the sentence - and yet they are felt as 'paraphrases' of the original sentence within the given context by native speakers.

A hierarchic systematization of paraphrases

As a result of the experiment mentioned above we have obtained a hierarchic systematization of paraphrasing means of Slovak. This systematization served as the basis for the experimental 'paraphrasing engine' of Slovak. From a deep structure, the paraphraser generates families of kindred surface forms in a cascade manner. The contents (the SRL-net) undergoes seven processes until it is changed into the surface shape:

1. Selection	What do you want to speak *about*
2. Mutation	
3. Figuration	*What* do you want to say about it
4. Domination	
5. Formation	
6. Focation	*How* do you want to say it
7. Intonation	

During **selection** a subset of nodes and arcs of the semantic net is chosen. You may want to speak about all entities and relationships participating in the story, but you may also choose only some of them, which you consider to be important. The semantic net may contain cycles as well. During *selection*, an acyclic subtree is chosen, which becomes a syntactic tree at the stage of formation.

For example *John's book is small* or *John walked and the book has been dedicated* would be **transselections** of (1).

During **mutation** the paraphraser is allowed to change the contents of the phrase, but within the range of elementary components of meaning only. One or more components of meaning may be eliminated or added to the semantic net. Resulting paraphrases will be spoken about as **transmuted.**

Among **transmutations** there are: synonymization *(bright → clear)*, antonymization *(cheap → expensive)*, magnonymization *(white → snowwhite, rain → downpour)*, minorization *(sweet → sweetish)*, deminutivization *(book → booklet)*, abstraction *(rose → flower)*, negativization *(dependent → independent, colored → colorless, to be → not to be)*, pluralization *(man → men)*, collectivization *(leaf → foliage)*, feminization *(actor → actress)*, aspectization *(to do → to have done)* and many others.

At the stage of **figuration**, basic semantic **figures** arise. **Transfigurated** sentences differ in meaning, but intersect in contents.[2] They stand for two different

[2] Except for the changes of content which arise by means of topic-focus articulation. Changes of meaning caused by word-reordering are generated at the stage of focation. This is an inconsistency, which has not been solved in our system.

semantic perspectives, functions, how the speaker looks upon one and the same fact. Typical representatives of transfigurations are perspectivization and conversion: *Leaves are falling from the tree → A leafy tree is falling off. Paul lent a book to Peter. → Peter borrowed a book from Paul. Paganini moved the audience by his beautiful violin play. → Paganini's play moved the audience by its beauty.*

Domination arranges the sentence vertically. **Transdominated** paraphrases shall differ in the choice of the central member of the sentence. Each node in an SRL-net may be chosen into the position of the central member of the future sentence or phrase. Remaining nodes shall then be syntactically subordinated to the central one: *three dwarfs → a triple of dwarfs, severe father → father's severity, John eats standing → John stands eating → John, who stands and eats.*

It is only during **formation**, that the amorphic SRL-nodes become a definite word-class (noun, verb, adjective...) and some of the grammatical cases are chosen to express semantic cases at the arcs of the semantic net. **Transformed** paraphrases (here in the narrow sense of the word) are different forms of expression of the same meaning.

The most common syntactic transformations are passivization (*Mary reads a book → The book is read by Mary)*, transitivization *(John stands and eats → John eats standing)*, infinitivization *(I came so as to see → I came to see)*, subordination *(speaking man → the man who speaks)*, coordination *(Ann can sing. Ann can dance. → Ann can sing and dance)*, apposition, complementization etc.

Another big group are condensation transformations. Among them there are: left and right nominalization *(John arrives → John's arrival → arrival of John)*, derivation *(wandering man → a wanderer, the place for a bell → belfry)* and composition *(with golden wings → goldenwinged, falling water → waterfall, lord of the land → landlord)*.

Focation arranges the sentence horizontally. **Transfocated** paraphrases differ in word-order. Actants may move from the topic-part of the sentence into the focus and vice versa. Word-order plays an important role in Slovak:

Janko nedaroval Marienke kvety.	(John did not present Mary with flowers.)
↓	
Janko nedaroval kvety Marienke.	(It was not Mary, to whom John gave the flowers.)
↓	
Marienke nedaroval kvety Janko.	(It was not John, who gave the flowers to Mary.)

Intonation changes the acoustic arrangement of the sentence. Different parts of the sentence may be stressed tonally within **transintonated** paraphrases: *John* is leaving → John is *leaving*.

The difference between transformation, transfiguration and transmutation is, that transformation alters the form, transfiguration alters the meaning and transmutation alters the contents. Transformed paraphrases differ in syntactic cases, but

have the same meaning. Transfigurated paraphrases differ in semantic cases, but have the same contents. Transmuted paraphrases differ in factual relationships (cases) themselves, they differ in contents.

An analogy with metallurgy

The process of transition from contents to form resembles the process of casting, forging and surface work with metals.

Selection means the choice of material. Mutation resembles putting admixtures into steel, which influences the final qualities of the material. After figuration (casting) you may bend the red-hot metal, but it is given *what* the product shall be in essence. Bending corresponds to domination. Formation corresponds to the eventual cooling off, when the product gets its final form. Focation and intonation are like grinding off and painting.

Computation of semantic relevance

Within the frame of the formal system described above, SAPFO is able to ascretain semantic relationships between Slovak phrases. Being given two phrases L_1 and L_2, the system tries to eliminate more and more semantic differences (components of meaning) between L_1 and L_2, until their meanings would intersect.

Formally: if the computer succeeds to find two sequences of synsemantic transformations $T_{11}, T_{12}, \ldots, T_{1n}$ and $T_{21}, T_{22}, \ldots, T_{2m}$ such that:

$$T_{11}T_{12}\ldots T_{1n}L_1 = L_1'$$
$$T_{21}T_{22}\ldots T_{2m}L_2 = L_2'$$

and L_1', L_2' intersect (they contain subparts unifiable by means of simple pattern matching):

$$L_1' \cap L_2' \neq \emptyset$$

then L_1 and L_2 *may* be semantically relevant. The sequence $T_{2m}^{-1}\ldots T_{21}^{-1}T_{11}\ldots T_{1n}$ is the semantic bridge between L_1 and L_2, and the semantic difference between L_1 and L_2 is given by the list of components of meaning associated with the transfomations being used.

In this way, SAPFO provides also a certain kind of question answering or information retrieval from a natural language database. An example:

Suppose, there is a query

Q = Nakupoval Jano cenné veci?
(Was John buying in precious things?)

and the database D contains (among others) the sentence

A = Pavol vedel o Petrovom predaji diamantu Janovi.
(Paul knew about Peter's sale of the diamond to John.)

After transformation of **Q** and **A** we get:

Pavol vedel o Petrovom predaji diamantu Janovi.		$/Nol^{-1}Nor^{-2}$
Pavol vedel, že Peter predal diamant Janovi.		$/Kon^{-1}$
Pavol vedel, že Jano kúpil diamant od Petra.		/Abs
Pavol vedel, že Jano kúpil drahokam od Petra.		$/Kdz^{-1}$
Pavol vedel, že ***Jano kúpil drahý kameň*** *od Petra.*	= ***A'***	
Nakupoval Jano cenné veci?		$/Que^{-1}$
Jano nakupoval cenné veci.		$/Reo^{-1}$
Jano kupoval cenné veci.		/Asp
Jano kúpil cenné veci.		/Der
Jano kúpil cennosti.		$/Plu^{-1}$
Jano kúpil cennosť.		/Syn
Jano kúpil drahokam.		$/Kdz^{-1}$
Jano kúpil drahý kameň.	= ***Q'***	

In English:

Paul knew about Peter's sale of the diamond to John.		$/Nol^{-1}Nor^{-2}$
Paul knew, that Peter sold the diamond to John.		$/Kon^{-1}$
Paul knew, that John bought the diamond from Peter.		/Abs
Paul knew, that John bought the jewel from Peter.		$/Kdz^{-1}$
Paul knew, that ***John bought the precious stone*** *from Peter.*	= ***A'***	
Was John buying in expensive things?		$/Que^{-1}$
John was buying in expensive things.		$/Reo^{-1}$
John was buying expensive things.		/Asp
John bought expensive things.		/Der
John bought valuables.		$/Plu^{-1}$
John bought a valuable.		/Syn
John bought a jewel.		$/Kdz^{-1}$
John bought a precious stone.	= ***Q'***	

We have

$$\frac{\begin{array}{l} Kdz^{-1}SynPlu^{-1}DerAspReo^{-1}Que^{-1}\mathbf{Q} = \mathbf{Q'} \\ Kdz^{-1}AbsKonNor^{-2}Nol^{-1}\mathbf{A} = \mathbf{A'} \\ \mathbf{Q'} \leq \mathbf{A'} \end{array}}{\mathbf{Q} \leq QueReoAsp^{-1}Der^{-1}PluSyn^{-1}KdzKdz^{-1}AbsKonNor^{-2}Nol^{-1}\mathbf{A}}$$

The sequence $ReoAsp^{-1}Der^{-1}PluSyn^{-1}AbsKonNor^{-2}Nol^{-1}$ is the semantic bridge between the meanings of **Q** and **A**. The members of the sequence say a lot about the kind and extent of differences between the query **Q** and answer **A**:

Nor^2, *Nol*, *Der* caused only formal changes. *Kon* changed the perspective of the sentence. Reo^{-1} eliminated the orientation of the event *buy/buy in* (+ex), *Asp* changed the aspect from progressive into finished (+perf), Plu^{-1} altered the number (-plu). *Abs* expanded the meaning of **A**; **A** becomes only a concrete instance of **A'** (**A** ≤ **A'**). *Syn* caused some changes within the range of lexical synonymy (**Q** ≈ **Q'**). Resulting semantic equation is:

$$\mathbf{Q} \approx \mathbf{A} \text{ +ex -perf -plu}$$

In this case, as we see, **Q** is found to be a perfectly relevant answer to **A**, still after application of 12 transformations.

Conclusion: A tool for linguistic analysis

The computer-paraphraser serves as a useful and inspiring tool for linguistic analysis. Suppose the linguist has a certain hypothesis about the functioning of some subsystem of language (syntax, semantics, word-formation, phraseology, stylistics etc.). Running the program and looking at the resulting paraphrases makes it evident how does the new rule or concept behave in the environment of other rules and concepts. Usually, the paraphrases are illustrative enough to show also, *in what way* the linguist's assumptions have been wrong.

Language is such a complex, manifold and vivid phenomenon, that no human linguist is able to formulate his hypotheses fully aware of all relationships in language, which can be relevant and influence the subject of his study. Human consciousness must always be narrowed by concentration.

Only a machine can store in memory and take into account all rules of all subsystems of language and test their behavior in mutual interaction. Only using a computer the researcher can test all hypotheses on large corpora and set himself free form the suggestive power of a particular example.

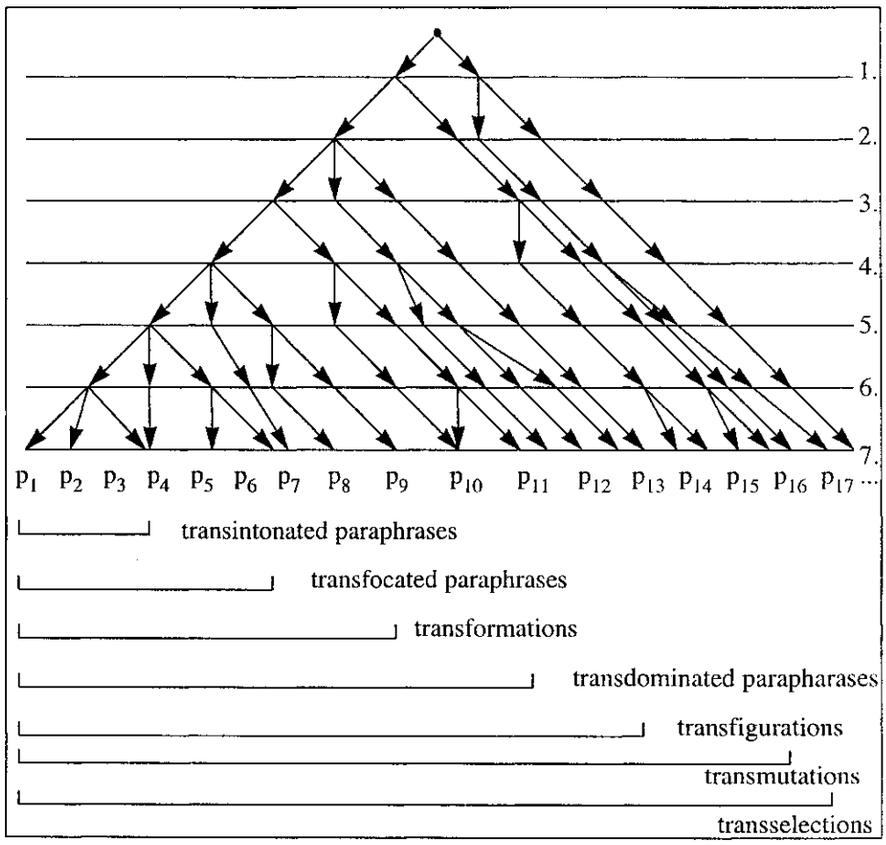


Fig. 2. The higher the level, where the paraphraser undeterministically undertakes some changes, the farther will be the form of resulting paraphrases.

RNDr. Emil Páleš, CSc.

Sapfo

PARAFRÁZOVAČ SLOVENČINY

počítačový nástroj
na modelovanie v jazykovede

Obálku navrhol Jozef Szabó
Výtvarný redaktor Jozef Szabó
Zodpovedná redaktorka PhDr. Jitka Madarásová

Vydanie prvé. Vydala VEDA, vydavateľstvo Slovenskej akadémie vied,
v Bratislave roku 1994 ako svoju 3074. publikáciu. AH 16,7 VH 17,26
Tlač SAP, spol. s r.o.

ISBN 80-224-0109-9

VEDA
VYDAVATEĽSTVO
SLOVENSKEJ
AKADÉMIE
VIED
BRATISLAVA
1994

ISBN 80-224-0109-9