

## Proces premeny obsahu na formu Sedemstupňový algoritmický model

Emil Páleš

Jazykovedný ústav L. Štúra SAV, Bratislava

### Počítačový model slovenčiny

SAPFO (Sense Apprehending Frame Operator) je počítačový model slovenského jazyka ako celku. Zahrňuje deväť jazykových podsystémov — hláskoslovie, tvaroslovie, slovtvorbu, skladbu, sémantiku, lexikálnu paradigmaticku, štylistiku, frazeológiu, asociatívne a konotatívne vazby — a umožňuje modelovať a pozorovať ich činnosť vo vzájomnej interakcii. Každý z týchto deviatich podsystémov jazyka je reprezentovaný jedným programovým modulom, programové moduly paralelne komunikujú a spolupracujú.

SAPFO pracuje ako parafrázovač slovenských viet. Používateľ napíše nejakú vetu alebo frázu a systém odpovedá generovaním všetkých takých slovenských viet a fráz, ktoré majú síce rozličné formy, ale sú pôvodnej vete významovo blízke. Na množine utvorených parafráz je potom bezprostredne viditeľné, do akej miery, resp. v akom smere počítač „pochopil“ či „nepochopil“ význam pôvodnej vety. Za základ pre počítačový parafrázovač poslúžil psycholingvistický experiment, pri ktorom sme ľudské subjekty požiadali, aby vlastnými slovami prerozprávali krátke rozprávkové príbehy. Vďaka tomuto experimentu sme vyvodili určité závery o tom, aké rozličné povrchové formy človek asociuje s tým-ktorým obsahom, a túto skúsenosť sme potom implementovali do počítača.

Typický výstup parafrázovača vyzerá napríklad takto

*Starec dal malej Anne žiarivý drahokam.*

*Anička dostala od starca žiarivý drahý kameň*

*Kameň, ktorý Anička dostala od starca, nebol lacný a žiaril*

*Žiara drahokamu, ktorý malá Anna dostala od starca ako dar*

*Starý muž obdaroval malú Annu kameňom. Bol drahý a žiaril*

*Starcov dar Anne — žiarivý kameň veľkej ceny*

*Darca Ankinho drahokamu bol starý*

*\*Anna obdržala od starého drahý žiarič*

*\*Starcov kamenný darček bol rozžiarený*

Úplný opis systému sme publikovali v inej práci (Páleš, 1993c)

### Reprezentácia významu

Systém SAPFO pracuje tak, že povrchové vetné formy najprv transformuje do hĺbkových významových štruktúr a z nich potom generuje ďalšie povrchové formy aplikáciou rozmanitých vyjadrovacích a parafrázovacích prostriedkov. Hĺbková reprezentácia má charakter stavebnice pozostávajúcej z troch druhov tehál: z nerozložiteľných plnovýznamových slovných koreňov (slovník SAPFO ich obsahuje 300), zo sémantických pádov (88) a z komponentov významu (66). Z nich možno kombinatoricky konštruovať veľké množstvo štruktúr, tzv. sémantických sietí. Napríklad nasledujúca sieť reprezentuje vetu *Anička dostala od starca žiarivý drahokam* a jej parafrázy:

da-/darov-/dost-<sup>+pr,tri</sup> +pret  
REC» Ann-<sup>+dem</sup>  
DON» osob-  
ATR» star-  
OBJ» kameň-  
ATR» drah<sub>1</sub>-  
«PDK žiar-<sup>+incl</sup> (1)

Program využíva pri svojej práci týchto deväť slovníkov: fonologický, morfológický, derivatologický, syntaktický, lexikálno-sémantický, synonymický a lexikálno-paradigmatický, frazeologický, štylistický a asociatívno-konotatívny.

### Hierarchická systematizácia a kaskádové generovanie parafráz

Obsah (reprezentovaný sieťou) prechádza počas formovania sa do viet (reprezentovaných lineárne) siedmimi základnými procesmi:

obsah	1. Selekcia	□	o čom chceme hovoriť
význam	2. Mutácia	]	čo o tom chceme povedať
	3. Figurácia		
forma	4. Dominácia	]	ako to chceme povedať
	5. Formácia		
	6. Fokácia		
	7. Intonácia		

Na každej z týchto siedmich úrovní možno počas procesu generovania modifikovať určité parametre, ktoré ovplyvnia výsledný vzhľad parafrázy. Čím na hlbšej úrovni sa niektorý parameter pozmení, tým vzdialenejšie povrchové formy parafráz vznikajú.