

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, slovotvorbu, skladbu, sémantiku, lexikálnu paradigmaticku, štylistiku, frazeológiu, asociatívne a konotatívne väzby — 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 spolu pracujú.

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-^{+perf +pret}
REC» Ann-^{+dem}
DON» osob-
ATR» star- (1)
OBJ» kameň-
ATR» drah₁-
«PDK žiar-^{+incl}

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

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
	2. Mutácia]	čo o tom chceme povedať
význam	3. Figurácia		
	4. Dominácia		
	5. Formácia]	ako to chceme povedať
	6. Fokácia		
forma	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.

Parafrázy, ktoré až do n-tej úrovne genézy šli spoločnou cestou, tvoria spoločnú rodinu n-tého stupňa. Napríklad transfigurácie tvoria rodinu takých parafráz, ktoré až do druhej genetickej úrovne majú spoločný pôvod a z tretej úrovne pokračovali rozličnou voľbou parametrov. Majú teda totožný obsah, ale nemusia mať rovnakú sémantickú perspektívu. Vetné transformácie a transpozície tvoria v tejto hierarchii špeciálnu podskupinu: je to rodina parafráz piateho stupňa. Tieto sa líšia gramatickou formou, no zhodujú významom.

Uvedené rodiny zároveň zodpovedajú hierarchickej typológii parafrázovacích prostriedkov, ktorú sme získali ako výsledok spomínaného psycholingvistického experimentu (Páleš, 1993a).

Parafrázovač môže vytvoriť jednu parafrázu na želanie — podľa zvolených parametrov — alebo kaskádovite generovať celé rodiny parafráz, pričom variabilné parametre sám cyklicky obmieňa.

Kaskádové generovanie parafráz

Pri **selekcii** sa zo sémantickej siete vyberá podmnožina tých hrán a vrcholov, o ktorých sme sa rozhodli hovoriť. Zvoliť môžeme všetky entity a vzťahy, ktoré sa príbehu zúčastňujú, resp. vybrať iba tie, ktoré pokladáme za dôležité z hľadiska komunikačného zámeru.

Napríklad *Anička dostala darček* a *Kameň žiaril* by boli **transselekcie** obsahu v sieti (1).

Počas **mutácie** je parafrázovaču dovolené zmeniť obsah frázy, avšak iba v rozsahu elementárnych komponentov významu (pozri Páleš 1993a alebo c). Jeden alebo viac komponentov významu možno v sieti vynechať, pripojiť či zmeniť. Vznikajú tým transmutované parafrázy.

Medzi **transmutácie** patria: synonymizácia (*priehľadný* → *priezračný*), antonymizácia (*drahý* → *lacný*), magnonymizácia (*biely* → *snehobiely*, *prší* → *leje*), minorizácia (*sladký* → *sladkastý*, *nasladlý*), deminutivizácia (*kniha* → *knižočka*), feminizácia (*herec* → *herečka*), pluralizácia (*človek* → *ľudia*), kolektivizácia (*list* → *lístie*), negativizácia (*farebný* → *bezfarebný*, *byť* → *nebyť*), aspektizácia (*robiť* → *urobiť*), frekventizácia (*chodiť* → *chodievať*), disperzia (*letieť* → *lietať*), dynamizácia (*je čierny* → *černie sa*), abstrakcia (*ruža* → *kvet*) a mnohé ďalšie.

V štádiu **figurácie** sa utvárajú základné sémantické figúry. **Transfigurované vety** sa líšia významovým stvárnením rovnakého obsahu. Zastupujú dve rozličné sémantické perspektívy, z akých sa hovoriaci pozerá na jeden a ten istý fakt: *Listy stromu opadávajú* → *Listnatý strom opadáva* → *Lístie padá zo stromu*; *Paganini uchvátil obecenstvo krásnou hrou na husle* → *Paganiniho hra na husle uchvátila obecenstvo svojou krásou* → *Krása*

Paganiniho hry uchvátila obecenstvo → *Paganiniho husle uchvátili obecenstvo*.

Typickými predstaviteľmi transfigurácií sú perspektivizácia (*Pekár vysypal múku* → *Pekárovi sa vysypala múka*) a konverzia (*Farby tvoria spektrum* → *Spektrum sa skladá z farieb*).

Pri **dominácii** sa veta štrukturuje vertikálne. **Transdominované** parafrázy sa líšia voľbou centrálného vetného člena. Ľubovoľný vrchol siete možno zvoliť za centrálny člen budúcej vety. Zvyšné členy sa mu potom syntakticky podriadia: *prísny otec* → *prísnosť otca*, *siedmi trpaslíci* → *sedmica trpaslíkov*, *splynený koks* → *koksový plyn*, *Jano vidí dievča čítajúce knihu* → *Janom videné dievča číta knihu*.

Až počas **formácie** sa amorfné slovné korene konštituuju do niektorého zo slovných druhov a významy prijímajú konkrétne gramatické formy. Sémantické pády sa zobrazia na niektorý z gramatických pádov. **Transformované** parafrázy (tu v úzkom zmysle slova) sú rozličné formy vyjadrenia toho istého významu.

Najbežnejšími syntaktickými transformáciami sú pasivizácia (*Katalyzátor urýchľuje reakciu* → *Reakcia je urýchľovaná katalyzátorom*), tranzitivizácia (*prichádza a spieva* → *prichádza spievajúc*), infinitivizácia (*Prídem, aby som sa pozrel* → *Prídem sa pozrieť*), participalizácia (*Jano vidí dievča. Dievča číta knihu* → *Jano vidí dievča čítajúce knihu*), komplementizácia (*Sestra sa vrátila domov. Sestra je tichá.* → *Sestra sa vrátila domov tichá*), apozícia (*Nitra je rieka. Nitra je modrá.* → *Rieka Nitra je modrá*), subordinácia (*učiteľ rozpráva* → *učiteľ, ktorý rozpráva*), koordinácia, adordinácia.

Druhú veľkú skupinu tvoria kondenzačné transformácie: Medzi ne patria ľavá a pravá nominalizácia (*Ján prišiel* → *Jánov príchod* → *príchod Jána*), derivácia (*putujúci muž* → *pútnik*, *miesto pre zvon* → *zvonica*, *veľmi zlatý* → *zlatučký*, *prestáť fajčiť* → *dofajčiť*), kompozícia (*so zlatými krídlami* → *zlatokrídly*, *muž, ktorý odlieva zvony* → *zvonolejec*), juxtapozícia, abreviácia, akronymizácia a ďalšie.

Fokácia utvára vetu horizontálne. **Transfokované** parafrázy sa líšia slovosledom a aktuálnym členením. Aktanty sa môžu pohybovať z rematickej do tematickej časti a naopak: *Janko nedaroval Marienke kvety* → *Janko nedaroval kvety Marienke* → *Marienke nedaroval kvety Janko*.

Intonácia mení zvukové stvárnenie vety. Dôraz na jednotlivé slová alebo časti slov je vyznačený kurzívou: *Jano prišiel* → *Jano prišiel*.

Analógia s metalurgiou

Proces premeny obsahu na formu sa podobá odlievaniu, kutiu a opracovávaníu 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, *čím* sa výsledný produkt stane. Ohýbaniu mäkkého kovu zodpovedá dominácia. Formácia znázorňuje kutie a úplné ochladenie kovu, ktorý tak nadobudne definitívny tvar. Nakoniec fokácia zodpovedá brúseniu a opracovávaní kovu za studena a intonácia povrchovej úprave, lešteniu, natieraníu.

Tekuté skupenstvo zodpovedá plnosti cítenia, ktorú nemôžeme vysloviť. Je tehotné množstvom foriem. Tuhé skupenstvo zodpovedá vyjadreniu, ktoré je však vždy už iba čiastočné.

Záver: počítačový nástroj v rukách jazykovedca

V rukách lingvistu poslúži počítačový model jazyka ako exaktný aplikačný a verifikačný nástroj, ktorý umožňuje skúmať a revidovať ľubovoľné jazykovedné hypotézy, myšlienky či teórie, a testovať ich správanie vo vzájomnej prepojenosti s ostatnými podsystémami a štruktúrami jazyka a to na veľkých korpusoch textov. Tým počítač pomáha jazykovedcovi—človeku oslobodiť sa od sugescie špeciálneho príkladu a preklenúť u človeka nevyhnutné obmedzenie vedomia koncentráciou.

Funkčnosť či nefunkčnosť parafrázovača je neúprosným objektívnym kritériom na posudzovanie adekvátnosti našich predstáv o fungovaní jazyka. Skúsenosť s počítačom pomáha jazykovedcovi dobrušovať základný jazykovedný pojmový aparát, ba často ho inšpiruje aj k netradičným postojom alebo úplne novým vhladom.

Literatúra

ALLEN, J.: Natural Language Understanding. Benjamin/Cummings, Menlo Park, 1987. 574 s.

HORECKÝ, J.: Ein System von Konzeptualisatoren für Künstliche Intelligenz. Proceedings of the Fourteenth International Congress of Linguistics, Berlin, 1987, s. 1187—1189.

KELEMEN, J.: Multiagent symbol systems and behavior-based robots. Applied Artificial Intelligence 7 (1993), s. 419—432.

PÁLEŠ, E.: Sémantické roly slovenských sloviess. Jazykovedný časopis 41, 1990, 1, s. 30—48.

PÁLEŠ, E.: Co-operation of Syntax and Semantics in Flexive Languages. Journal of Experimental & Theoretical Artificial Intelligence, 2, 1990, 1, s. 1—24.

PÁLEŠ, E.: Natural Language Paraphraser — A Computer Tool for Linguists. Computers & Artificial Intelligence 12, 1993a, 5, s. 491—510.

PÁLEŠ, E.: Slovtvorná parafráza na počítači. Celostné modelovanie jazyka. Jazykovedný časopis 44, 1993b, 2, s. 108—125.

PÁLEŠ, E.: Parafrázovač slovenčiny. Počítačový nástroj na modelovanie v jazykovede. Jazykovedný ústav L. Štúra SAV, Bratislava, 1993c. 250 s.

PANEVOVÁ, J.: Formy a funkce ve stavbě české věty. Academia, Praha, 1980. 224 s.

WINOGRAD, T. — FLORES, F.: Understanding Computers & Cognition. Addison—Wesley, Reading, 1987.