

červený drak

drak je červený	^
červený drak	^Nol
červená dračica	^FemNol
ohnivočervený drak	^MagNol
dračica červená ako oheň	^KomFemNor
drak s červeňou	^Nor
červeň draka	^Nom
farebné zvierá	^Abs ²

Všimnite si, že synsémantické transformácie menia tak význam ako aj formu. Podľa nášho názoru dve konštrukcie sú naozaj synonymné vtedy a len vtedy, ak sú totožné. Preto náš prístup je založený nie na štúdiu synonymných (Sgall a kol., 1986), ale skôr nesynonymných vzťahov v jazyku. Skúsenosť s ozajstnými komunikačnými situáciami ukazuje, že komunikujúci často používajú nesynonymné výrazy v úlohe synonymných vďaka tzv. neutralizácii sémantických dištinkcií v kontexte (Apresian, 1974). Tak napríklad slovo 'učiteľka' sa často očakáva ako odpoveď na otázku o 'učiteľovi' (aj keď nejde o synonymá) apod.

4. Návrat informácie z korpusu textov

Formulujeme problém návratu informácie v prirodzenom jazyku (Information retrieval in natural language) takto: Máme rozsiahly korpus textov **D** a otázku používateľa **Q**. Úlohou systému je vypísať z **D** tie vety, ktoré kandidujú byť relevantnou odpoveďou na otázku **Q**.

Schopnosť tvoriť parafrázy pomáha k čiastočnému riešeniu tohto problému. V zmysle úvah uvedených vyššie, definujeme za 'parafrázu' vety **S** ľubovoľnú vetu **S'**, ktorá vznikla z **S** aplikáciou postupnosti synsémantických transformácií pod kontrolou unifikačnej gramatiky

$$S' = T_1 T_2 \dots T_n S$$

Teda systém parafráz frázy **S** pozostáva nie len zo synonymných fráz, ale aj všetkých takých, ktorých význam má s významom pôvodnej frázy spoločný prienik alebo iný preddefinovaný vzťah. Ak sa podarí nájsť k otázke **Q** takú postupnosť transformácií $T_1, T_2, \dots, T_n$, že $T_1 T_2 \dots T_n Q = S$ a $S \in D$, pokladáme **S** za kandidáta na relevantnú odpoveď ku **Q** a vypíšeme **S** na výstup ako jednu z možných odpovedí. Postupnosť $T_1 T_2 \dots T_n$ tvorí 'sémantický most' medzi **Q** a **S**, t.j. definuje významový vzťah medzi **S** a **D** rozložený analyticky do jednoduchých posunov významu a formy 'sémantický rozdiel' medzi **S** a **Q** je daný sumou sémantických diferencií (príznakov) asociovaných s transformáciami, ktoré sme použili.

5. Príklad

Predpokladajme, že korpus **D** obsahuje (okrem iného) vetu

Pavol vedel o Petrovom predaji diamantu Janovi

a otázka znie

Nakupoval Jano cenné veci?

Použitím synsémantických transformácií pretransformujeme **Q** a **S** na tzv. kanonické tvary Q' , S' . Tieto sú potom unikovateľné pomocou jednoduchého testu na prienik

Pavol vedel o Petrovom predaji diamantu Janovi	/Nol ⁻¹ Nor ⁻²
Pavol vedel, že Peter predal diamant Janovi	/Kon ⁻¹
Pavol vedel, že Jano kupil diamant od Petra	/Abs
Pavol vedel, že Jano kúpil drahokam od Petra	/Kdz ¹
Pavol vedel, že Jano kúpil drahý kameň od Petra	- S'

Nakupoval Jano cenné veci ?	/Que ¹
Jano nakupoval cenné veci	/Reo ⁻¹
Jano kupoval cenné veci	/Asp
Jano kúpil cenné veci	/Der
Jano kupil cennosti	/Plu ⁻¹
Jano kúpil cennosť	/Syn
Jano kúpil drahokam	/Kdz ¹
Jano kúpil drahý kameň	= Q'

Vidíme, že Q' a S' unifikujú, postupnosť $\text{NolNor}^2\text{KonAbs}^1\text{SynPlu}^1\text{DerAspReo}^1$ tvorí sémantický most medzi Q' a S' a teda **S** je relevantnou odpoveďou na otázku **Q**.

Literatura

- ANICK, P. - PUSTEJOVSKY, J. An Application of Lexical Semantics to Knowledge Acquisition from Corpora. COLING '90.
- APRESIAN, J. D. Lexical Semantics. Synonymical Means of Language. Moscow, Nauka 1974. 368 p.
- PÁLEŠ, E. Co-operation of Syntax and Semantics in Flexible Languages. Journal of Experimental & Theoretical Artificial Intelligence, 2, 1990, No. 1, pp. 1-24.
- PUSTEJOVSKY, J. The Generative Lexicon. Computational Linguistics, 17, 1991, No. 4.
- SGALL, P. - HAJIČOVÁ, E. - PANEVOVA, J. The Meaning of the Sentence. Praha, Academia 1986.