3 Teoretická lingvistika a strojový překlad II

3.1 Porovnání závislostní a složkové reprezentace

závislostní reprezentace: prvořadé je vyjádření vztahu závislosti, úspornější, jednodušší sémantická interpretace, bližší významovému zápisu, hloubkové struktuře (výhodné pro transfer); vztah mezi textovým řetězcem a reprezentací není tak přímočarý jako u složkové (frázové) reprezentace

složková reprezentace: prvořadé je vyjádření vztahu těsnosti, bezprostřední vztah k textovému řetězci (CF-PSG); obtížnější sémantická interpretace, vztah k hloubkové struktuře není přímočarý

CF-PSG v čisté podobě, bez identifikace řídících členů (heads) složek se používá jen zřídka. Na druhé straně žije závislostní gramatika často v (implicitní) symbióze s CF-PSG. Příklad: kombinace CF-PSG se závislostní interpretací:

(1) Bezkontextová pravidla se závislostní interpretací

```
M  -> H            M    ->    H    N            M  ->   N   H
h     h            h          h    n            h       n   h
|     |           / \         |                / \          |
x     x          x   n        x               n   x         x
```

3.2 Strukturované kategorie místo atomických symbolů

pro řešení shody, valenčních rámců, (sémantickou, závislostní) interpretaci pravidel CF-PSG

(2) Pravidla se strukturovanými kategoriemi: shoda a valence

```
S   ->  NP           VP
        number:X     number:X
        gender:Y     gender:Y
        person:Z     person:Z
        case:nom

VP      V            NP
        case:X       case:X
```

3.3 Sestavy rysů (‘feature structures’)

Tzv. distinktivní rysy užívali pro popis vlastností lingvistických objektů již strukturalisté. Sestavy rysů jako základní nástroj popisu jazyka (lexikálních položek, typů a výskytů konstrukcí jazyka) se začaly používat v unifikačních gramatikách od konce 70. let. Sestava rysů je množina dvojic atribut-hodnota, přičemž hodnota může být atomická nebo komplexní, t.j. další sestava rysů.

(3) Reprezentace lexikálních položek pomocí sestav rysů

$$\begin{bmatrix} \text{CATEGORY} & \textit{np} \\ \text{CASE} & \left\{\textit{nom, acc, voc}\right\} \\ \text{LEXEME} & \textit{kůň} \\ \text{FORM} & \textit{koně} \\ \text{NUMBER} & \textit{pl} \\ \text{PERSON} & \textit{3} \\ \text{GENDER} & \textit{ma} \end{bmatrix} \qquad \begin{bmatrix} \text{CATEGORY} & \textit{vp} \\ \text{VFORM} & \textit{finite} \\ \text{LEXEME} & \textit{zařehtat} \\ \text{FORM} & \textit{zařehtali} \\ \text{SUBJECT} & \begin{bmatrix} \text{CATEGORY} & \textit{np} \\ \text{CASE} & \textit{nom} \\ \text{NUMBER} & \textit{pl} \\ \text{PERSON} & \textit{3} \\ \text{GENDER} & \textit{ma} \end{bmatrix} \end{bmatrix}$$

3.4 Unifikace

operace, která vytváří sjednocení navzájem si neodporujících specifikací

(4) Shoda pomocí sestav rysů a unifikace

$$
\begin{bmatrix}
\text{CATEGORY} & \textit{np} \\
\text{CASE} & \textit{nom} \\
\text{LEXEME} & \textit{ovce} \\
\text{FORM} & \textit{ovce} \\
\text{NUMBER} & \_ \\
\text{PERSON} & \textit{3} \\
\text{GENDER} & \textit{f}
\end{bmatrix}
\quad
\begin{bmatrix}
\text{CATEGORY} & \textit{vp} \\
\text{VFORM} & \textit{finite} \\
\text{LEXEME} & \textit{zabečet} \\
\text{FORM} & \textit{zabečely} \\
\text{SUBJECT} & \begin{bmatrix}
\text{CATEGORY} & \textit{np} \\
\text{CASE} & \textit{nom} \\
\text{NUMBER} & \textit{pl} \\
\text{PERSON} & \textit{3} \\
\text{GENDER} & \left\{\textit{ma,f}\right\}
\end{bmatrix}
\end{bmatrix}
$$

```
     S                  -->     NP               VP
category = sentence      category = np     category = vp
                                           subject = NP
                                           vform = finite
```

(5) Tzv. *head features*

$$
\begin{bmatrix}
\text{CATEGORY} & \textit{np} \\
\text{HEAD} & \begin{bmatrix}
\text{CASE} & \textit{nom} \\
\text{LEXEME} & \textit{ovce} \\
\text{FORM} & \textit{ovce} \\
\text{NUMBER} & \_ \\
\text{PERSON} & \textit{3} \\
\text{GENDER} & \textit{feminine}
\end{bmatrix}
\end{bmatrix}
\quad
\begin{bmatrix}
\text{CATEGORY} & \textit{vp} \\
\text{HEAD} & \begin{bmatrix}
\text{VFORM} & \textit{finite} \\
\text{LEXEME} & \textit{zabečet} \\
\text{FORM} & \textit{zabečely} \\
\text{SUBJ} & \begin{bmatrix}
\text{CATEGORY} & \textit{np} \\
\text{HEAD} & \begin{bmatrix}
\text{CASE} & \textit{nom} \\
\text{NUMBER} & \textit{pl} \\
\text{PERSON} & \textit{3} \\
\text{GENDER} & \left\{\textit{ma, f}\right\}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

```
     S     -->     NP     VP
                               <VP category> = vp
                               <VP head subj> = NP
                               <VP head vform> = finite
                               <S category> = sentence
                               <S head> = <VP head>
```

$$
\begin{bmatrix}
\text{CATEGORY} & \textit{sentence} \\
\text{HEAD} & \begin{bmatrix}
\text{VFORM} & \textit{finite} \\
\text{LEXEME} & \textit{zabečet} \\
\text{FORM} & \textit{zabečely} \\
\text{SUBJ} & \begin{bmatrix}
\text{CATEGORY} & \textit{np} \\
\text{HEAD} & \begin{bmatrix}
\text{CASE} & \textit{nominative} \\
\text{LEXEME} & \textit{ovce} \\
\text{FORM} & \textit{ovce} \\
\text{NUMBER} & \textit{plural} \\
\text{PERSON} & \textit{3} \\
\text{GENDER} & \textit{feminine}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

3.5 Důvody pro rozšíření nekontextové frázové gramatiky

Pro některé jevy jazyka není obyčejná CF-PSG explanatorně ani deskriptivně adekvátní,[1] např. dává odlišnou reprezentaci pro synonymní činné a trpné věty. CF-PSG se proto někdy používá pro generování **hloubkové struktury**, která se transformuje na **povrchovou strukturu**.

Reprezentace synonymních výrazů: *Peníze rozdělilo ministerstvo.* vs. *Peníze byly rozděleny ministerstvem.*

V "obyčejné" CF-PSG se těmto synonymním větám přiřadí různé struktury.

Řešení:

- transformace 'deep structure' na 'surface structure', tedy více rovin
- lexikalizace gramatiky, tedy více rovin, ale bez transformace; podoba 'deep structure' pak už nezávisí na derivační historii (výhodné pro jazyky s volným slovosledem); gramatika zůstaává formálně bezkontextová, umožňuje parciální interpretace, je nezávislá na směru zpracování (deklarativní)

Reprezentace homonymních výrazů: *kritika našeho premiéra*

- strukturně s atomickými kategoriemi
- lexikálně s komplexními kategoriemi

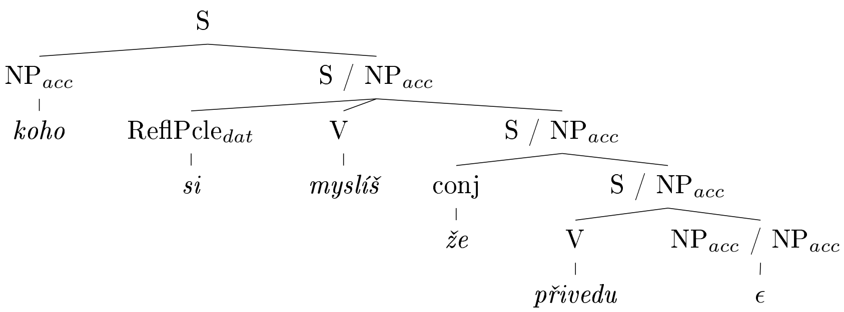
Některé jevy lze reprezentovat jednoznačně už na povrchové rovině: *Procházel se v parku s pivoňkami.*

Reprezentace nevyjádřených větných členů: *Honza mi slíbil zařídit audienci.* vs. *Honza mi přikázal zařídit audienci.* — *Včera jsem potkal Natálii. Už od školy pracuje v BISu.* — *Já rád vdolky a ty holky.*

Reprezentace tzv. neprojektivních konstrukcí stromovým grafem: *Plot už Pavel nestihl natřít* — *Koho si myslíš, že jsem včera potkal?* — *Panenku Barbie dostávají i menší děti než Mařenka.* — *Boris se odmítá oholit.*

Transformace jsou ovšem příliš mocný nástroj (cokoli lze přesunout kamkoli, problém výběru toho pravého transformačního pravidla při analýze). Sémantická interpretace nemůže také vycházet jen z hloubkové struktury v Chomského smyslu (slovosled je důležitý např. pro stanovení dosahu kvantifikátorů).
Rozdíl mezi hloubkovou a povrchovou reprezentací není podmíněn transformacemi. Všechny uvedené případy lze řešit pomocí strukturovanějšího systému kategorií (viz závislostní interpretace pravidel CF-PSG výše).

(6) Řešení neprojektivních konstrukcí pomocí komplexních kategorií



[1] Různé druhy 'adekvátnosti':

- empiricky adekvátní (gramatika umí rozhodnout, zda řetězec patří do jazyka)
- deskriptivně adekvátní (gramatika přiřazuje řetězcům adekvátní struktury)
- explanatorně adekvátní (gramatika odvozuje jazykové jevy z obecnějších principů)

3.6 Transfer

Pravidla transferu lze rozdělit na:

- lexikální
- strukturní: obecná nebo specifická (jejichž aplikace závisí např. na morfologických kategoriích nebo na lexikálním obsazení prvků struktury)

Pravidla transferu mohou být užita na:

- stromovém grafu
- sestavě rysů
- souboru kategorií popisujících nejmenší jednotky roviny

Pravidla transferu mohou mít přístup:

- pouze k části reprezentace (např. k hloubkové struktuře)
- ke všem údajům reprezentace (včetně povrchového slovosledu, morfologických kategorií)

Při aplikaci pravidel je třeba dodržovat určitý ppostup, který respektuje:

- rekursivní vlastnosti representace (výrazu jazyka): transfer postupuje od nejvyšší úrovně struktury k nižším úrovním
- míru obecnosti pravidel transferu: specifičtější pravidla se uplatňují přednostně

LITERATURA:

Eva Hajičová a kol. (red.) Studie z transformační gramatiky I - III, skripta MFF UK, SPN 1980, Praha

Peter Sells Lectures on Contemporary Syntactic Theories, CSLI Lecture Notes 3, CSLI, Stanford, USA

Stuart M. Shieber An Introduction to Unification-based Approaches to Grammar, CSLI Lecture Notes 4, CSLI, Stanford, USA

Carl Pollard & Ivan Sag Head-Driven Phrase Structure Grammar, University Of Chicago Press, 1994

Chomsky, N. (1982) Lectures on Government and Binding. Foris, Dordrecht.

Bresnan, J. (1982) (eds.) The mental representation of grammatical relations. MIT Press, Cambridge, Mass., USA.

Oehrle, R.T. et al. (1988) (eds.) Categorial Grammars and natural language structures. Reidel, Dordrecht.

Gazdar, G. et al. (1985) Generalized phrase structure grammar. Blackwell, Oxford.