

# Závislostní redukční analýza přirozených jazyků\*

Markéta Lopatková,<sup>1</sup> Martin Plátek<sup>2</sup> a Vladislav Kuboň<sup>3</sup>

<sup>1</sup> CKL MFF UK, Praha  
lopatkova@ckl.mff.cuni.cz

<sup>2</sup> KTIML MFF UK, Praha  
platek@ksi.ms.mff.cuni.cz

<sup>3</sup> ÚFAL MFF UK, Praha  
vk@ufal.mff.cuni.cz

**Abstrakt** V tomto článku vyložíme podstatu závislostní redukční analýzy (DAR, dependency analysis by reduction) a její souvislost s pojmy závislost a závislostní strom. Výklad budeme ilustrovat příklady z češtiny, což je jazyk s (výrazně) volným slovosledem. Tento výklad shrnuje základní rysy vývojových postupů závislostní syntaxe. Bude využit jako podklad pro ověřování (a vysvětlování) adekvátnosti formálních a počítačových modelů těchto postupů.

## 1 Slovo úvodem k redukční analýze

Větná stavba angličtiny a dalších jazyků s pevným slovosledem se formálně popisuje frázovými gramatikami. Popisy větné stavby latiny, italštiny, němčiny, slovanských jazyků, arabštiny a dalších jazyků bývají častěji a adekvátněji založeny na postupech, které se souhrnně nazývají závislostními. Oba principy lze vykládat pomocí pozorování postupného zjednodušování jednotlivých vět jazyka, pomocí tzv. redukční analýzy. Principy frázové redukční analýzy a závislostní redukční analýzy se však podstatně liší. Frázovou redukční analýzu (jazyků s pevným slovosledem) lze přímo modelovat analýzou zdola pomocí frázových (Chomského) gramatik. Na základě tohoto textu by čtenář měl nahlédnout, že závislostní redukční analýzu jazyků s volným slovosledem je nutné modelovat podstatně jinak.

Abychom přiblížili význam předkládaného textu, naznačíme, v čem se syntaktická analýza (větný rozbor) vět přirozeného jazyka liší od analýzy vět (textů) umělého (např. programovacího) jazyka. U umělých jazyků určujeme stavbu vstupní věty, přičemž obvykle máme k dispozici formálně zapsanou bezkontextovou gramatiku příslušného jazyka (obvykle vzniká gramatika zároveň s jazykem), navíc běžně požadujeme gramatiku takovou, která zaručuje jednoznačnost rozboru každé věty jazyka.<sup>4</sup>

\* Tento článek je výsledkem výzkumu, který probíhá za podpory grantu MŠMT LN00A063 a grantu GA ČR No. 201/02/1456

<sup>4</sup> Ve formálních jazycích se obvykle mluví o ‚symbolech abecedy‘ a ‚slovech‘, my budeme nadále používat u formálních i přirozených jazyků termínů ‚slovo (ze slovníku)‘ a ‚věta‘.

U přirozených jazyků je situace odlišná – při větném rozboru také hledáme stavbu vstupní věty. Ta však nebývá určena jednoznačně. Místo ucelené formální gramatiky máme k dispozici dovednost větného rozboru – učíme se jej ve školách, je popisován (často pomocí implicitních pravidel) v mluvnicích daného jazyka.

Mluvnice češtiny předpokládají, že člověk rozumí smyslu rozebírané věty dříve, než začne větný rozbor provádět (citujme zde Učebnici větného rozboru [10]: „Správný rozbor větný není možný bez přesného porozumění větě, ... “). Automatická syntaktická analýza (podle formální gramatiky) porozumění větě nepředpokládá, ani je nemá k dispozici. Je naopak jednou z prvních fází při počítačovém modelování smyslu věty.

Jaký je vztah větného rozboru k redukční analýze? Jednoduše řečeno, větný rozbor češtiny je založen na elementárnější schopnosti provádět redukční analýzu, tj. postupně zjednodušovat zkoumané věty. Cestu od větného rozboru k počítačové syntaktické analýze vybraného jazyka lze rozdělit na následující kroky:

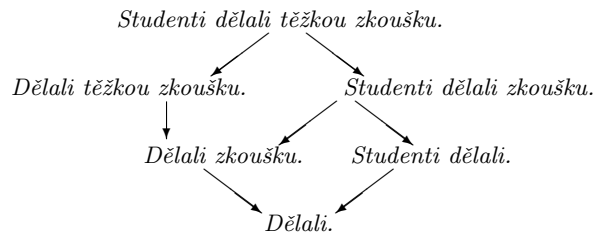
- (a) vyhledání a popis pravidel redukční analýzy jazyka,
- (b) transformace pravidel redukční analýzy do pravidel a omezení závislostního formalismu,
- (c) sestrojení (vybrání) vhodného analyzátoru pro daný soubor závislostních pravidel a omezení.

V tomto článku se budeme věnovat pozorováním patřícím bodu (a) a částečně bodu (b), se zaměřením na závislostní analýzu češtiny.

Redukční analýzou se zde zabýváme především proto, abychom získali jasnou představu o tom, jak ji formálně a počítačově modelovat. K tomuto účelu jsou zavedeny a studovány restartovací automaty (viz např. [3]).

Následující příklad ilustruje zjednodušeně metodiku závislostní redukční analýzy.

**Příklad 1.** Větu „*Studenti dělali těžkou zkoušku.*“ lze při zachování syntaktické správnosti zjednodušit dvěma způsoby (viz též schéma na obrázku 1) – vypuštěním slova *studenti*, nebo vypuštěním slova *těžkou* (ale už ne vypuštěním slova *zkoušku* – věta „*\*Studenti dělali těžkou.*“ není správně utvořená). Ve druhém kroku můžeme vypustit slovo *těžkou* (v první větvi analýzy), nebo slovo *studenti*, či slovo *zkoušku* (ve druhé větvi). V posledním kroku lze vypustit slovo *zkoušku* (v první větvi), nebo slovo *studenti*.



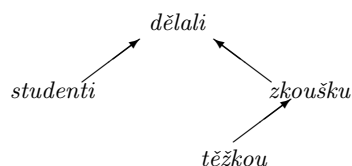
**Obrázek 1.** Schéma DAR pro větu „*Studenti dělali těžkou zkoušku.*“

Schéma DAR úzce souvisí se závislostním stromem (na obrázku 2 je závislostní strom pro větu „*Studenti dělali těžkou zkoušku.*“).

(i) Skutečnost, že nějaké slovo lze z věty vypustit tak, že získáme větu jednodušší, znamená, že závisí na některém slově ze zkrácené věty (rozvíjí ho).

(ii) Dvě slova lze postupně vypustit v libovolném pořadí, právě když jsou vzájemně nezávislá.

(iii) Navíc se ukazuje, že některá slovní spojení (např. spojení předložky s podstatným jménem; viz např. příklad 4 dále) je nutno vypouštět v jednom kroku – i v tomto případě je někdy vhodné určovat závislosti.



**Obrázek 2.** Závislostní strom věty „Studenti dělali těžkou zkoušku.“

Předchozí příklad ilustruje postup, jak pomocí DAR získat ve větě informaci o závislostech (vztazích mezi rozvíjenými a rozvíjejícími členy věty). Můžeme si všimnout, že vezmeme-li věty „Těžkou zkoušku studenti dělali.“ či „Těžkou dělali studenti zkoušku.“, nebo jejich další permutace, získáme zcela analogická schémata redukce jako pro větu původní, tedy vypuštěná slova (slovní spojení) v jednotlivých redukčních krocích budou stejná. Závislostní redukční analýza nám tímto způsobem umožňuje pozorovat, do jaké míry jsou závislosti invariantní vůči slovosledu.

Následující kapitoly se podrobněji zabývají závislostní redukční analýzou a určováním závislostí na základě redukční analýzy.

## 2 Závislostní redukční analýza

**Závislostní redukční analýza (DAR)** spočívá v postupném zjednodušování věty – každý krok DAR je reprezentován právě jednou **operací redukce**, která může být realizována dvěma způsoby:

(i) vypuštěním alespoň jednoho slova vstupní věty, nebo

(ii) nahrazením (obecně nesouvislého) podřetězce věty kratším podřetězcem.

Možnost aplikovat určitou redukci je podmíněna zachováním některých (alespoň prvního) z následujících **principů DAR**:

(a) zachování správnosti stavby věty (syntaktická správnost);

(b) zachování lemmatu (slovníkového hesla) a vybrané morfologické značky (soubor morfologických kategorií, které charakterizují daný výskyt slova);

(c) zachování významu původních slov ve větě (reprezentován např. valenčním rámcem,<sup>5</sup> či vhodným ekvivalentem v jiném jazyce);

(d) zachování významové samostatnosti věty (významově samostatná věta je taková věta, která vyslovena samostatně nevyvolává nutně další otázky).<sup>6</sup>

<sup>5</sup> Valenční rámec, který popisuje syntakticko-sémantické vlastnosti slova, odpovídá jednomu jeho významu, více viz např. [6].

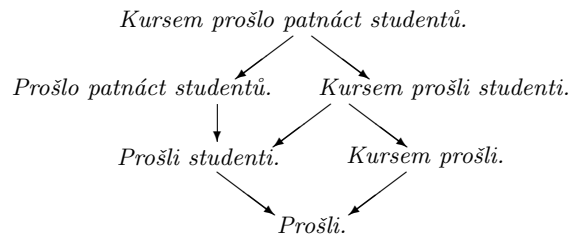
<sup>6</sup> V lingvistické terminologii to znamená, že významově samostatná věta se skládá ze slovesa, všech jeho sémanticky ‚povinných‘ doplňků a (rekurzivně) jejich ‚povinných‘ doplňků, přičemž ‚povinná‘ doplňky jsou ta doplňky, která musí mluvčí i posluchač znát pro dobré porozumění větě, více viz [7].

V závislosti na konkrétním úkolu (např. kontrola gramatické správnosti) lze tyto požadavky na DAR uvolňovat; principy, které nejsou uvolněny, potom označujeme jako **platné principy DAR** (např. v příkladu 1 byl uvolněn požadavek zachování významové samostatnosti věty – původní věta byla redukována až na jednočlennou větu ‚*Dělali.*‘, což už není samostatná věta).

Pokud lze v určitém kroku DAR aplikovat určitou redukci při zachování platných principů, mluvíme o **přípustné redukci**. Pomocí všech přípustných redukcí se lze dostat ke všem **přípustným zjednodušením** zpracovávané věty. Původní, nezjednodušenou větu budeme (z technických důvodů) považovat též za přípustné zjednodušení věty.

**Schématem DAR (redukčním schématem) věty jazyka** nazveme orientovaný graf, jehož uzly reprezentují všechna přípustná zjednodušení dané věty (včetně původní věty) a jehož hrany odpovídají všem přípustným redukcím, které lze aplikovat vždy na počáteční uzel hrany a jejichž výsledkem je přípustné zjednodušení věty v koncovém uzlu hrany.

**Příklad 2.** Redukční schéma věty ‚*Studenti dělali těžkou zkoušku.*‘ z příkladu 1 ilustruje redukce typu (i) – v každém kroku DAR je vypuštěno jedno slovo vstupní věty, přičemž možnost větvení zachycuje nedeterministickou povahu DAR. Redukce typu (ii) je ilustrována na možném zjednodušování věty ‚*Kursem prošlo patnáct studentů.*‘, jejíž redukční schéma je na obrázku 3. První krok první větve schématu se realizuje vypuštěním slova *kursem* (podle bodu (i), opět je uvolněn požadavek (d) na zachování významové úplnosti věty), v prvním kroku druhé větve je řetězec *prošlo patnáct studentů* nahrazen kratším řetězcem *prošli studenti* (podle bodu (ii)). Dále redukce pokračuje obdobně.



**Obrázek 3.** Schéma DAR pro větu ‚*Kursem prošlo patnáct studentů.*‘

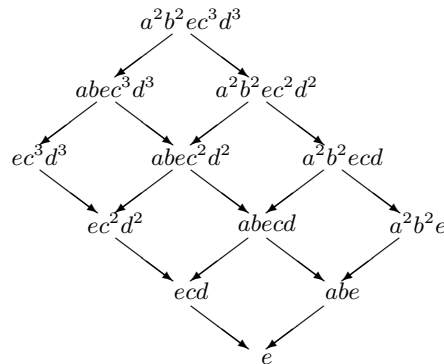
### 3 Redukční struktura a závislostní strom

Schéma DAR umožňuje zavést a klasifikovat různé typy vztahů. Na základě takto zavedených vztahů definujeme redukční strukturu věty. Podíváme se na její vztah k závislostnímu stromu.

Mějme jazyk  $L$ , větu  $v \in L$ ,  $v = v_1v_2\dots v_m$ , kde  $v_1, v_2, \dots, v_m$  jsou slova, a schéma DAR věty  $v$ . Řekneme, že slova  $v_i$ ,  $i \in N$ ,  $N \subseteq \{1, 2, \dots, m\}$  tvoří **redukční komponentu**, pokud jsou všechna  $v_i$  vypouštěna vždy najednou (tj. v redukčním schématu jsou všechna  $v_i$  vypouštěna vždy v jednom kroku, kterému odpovídá jedna hrana schématu). Řekneme, že slovo  $v_i$  je (**redukčně**) **závislé** (příp. **v redukci závislé**) na slově  $v_j$ , pokud je ve všech větvích DAR slovo  $v_i$

vypuštěno dříve než  $v_j$ ; slovo  $v_j$  budeme nazývat slovem (**v redukci**) **řídícím**. Řekneme, že slova  $v_i$  a  $v_j$  jsou **redukčně nezávislá**, pokud jsou vypouštěna v libovolném pořadí (tj. existuje větev DAR, ve které je slovo  $v_i$  vypuštěno před slovem  $v_j$ , a existuje větev DAR, ve které je slovo  $v_j$  vypuštěno před slovem  $v_i$ ).

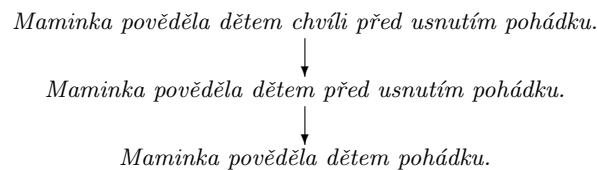
**Příklad 3.** Mějme formální jazyk  $L = \{a^i b^i e c^n d^n \mid i, n \geq 0\}$ . Na obrázku 4 je schéma redukční analýzy věty  $a^2 b^2 e c^3 d^3$ . Přijmeme pro jednoduchost výkladu předpoklad, že vypouštět v jednom kroku se smějí pouze dvojice nejbližších slov, pokud je více možností. Potom v této větě a větách redukovaných tvoří sousední slova  $a$  a  $b$  redukční komponentu, podobně sousední slova  $c$  a  $d$  tvoří také redukční komponentu; slova  $a, b, c$  a  $d$  jsou redukčně závislá na slově  $e$ ; slova  $a$  a  $b$  jsou redukčně nezávislá na slovech  $c$  a  $d$ . Jazyk  $L$  je příkladem jazyka s pevným slovosledem. Jazyk s volným slovosledem a ‚stejnými‘ redukcemi a (ne)závislostmi získáme tak, že budeme uvažovat množinu všech permutací vět z  $L$ .



**Obrázek 4.** Schéma DAR pro větu  $a^2 b^2 e c^3 d^3$

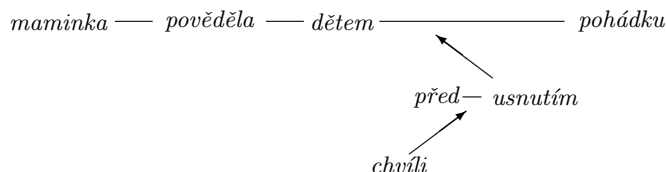
Na základě zavedených vztahů redukční závislosti a redukční komponenty definujeme redukční strukturu věty – ilustrujeme ji na následujícím příkladu.

**Příklad 4.** Redukční schéma věty ‚Maminka pověděla dětem chvíli před usnutím pohádku.‘, které zachovává všechny principy DAR z předchozí sekce (tj. včetně principu (d) zachování významové samostatnosti věty), je na obrázku 5 – předložková skupina *před usnutím* musí být redukována v jednom kroku; sloveso *povědět* má tři ‚povinná‘ doplňení, která odpovídají podmětu a tzv. přímému a nepřímému předmětu, podstatné jméno *maminka* nemá povinná doplňení, proto má významově samostatná věta tvar ‚Maminka pověděla dětem pohádku.‘.



**Obrázek 5.** Schéma DAR pro větu ‚Maminka pověděla dětem chvíli před usnutím pohádku.‘ při uplatnění principu zachování významové samostatnosti věty.

**Redukční strukturu** budeme zachycovat diagramem, ve kterém uzly reprezentují jednotlivá slova věty, horizontální hrany spojují slova tvořící redukční komponentu (hranou jsou spojeny vždy dvě sousední slova redukční komponenty).<sup>7</sup> Šikmé hrany odpovídají redukčním závislostem; považujeme je za orientované od závislého slova (resp. od celé redukční komponenty) ke slovu řídicímu (resp. opět k celé redukční komponentě, pokud je pro dané slovo / komponentu řídicí). Lineární uspořádání uzlů (zleva doprava) zachycuje slovosled (pořadí slov ve větě). Na obrázku 6 je redukční struktura reprezentující větu „*Maminka pověděla dětem chvíli před usnutím pohádku.*“



**Obrázek 6.** Redukční struktura pro větu „*Maminka pověděla dětem chvíli před usnutím pohádku.*“

Stavba (české) věty se tradičně zachycuje závislostním stromem. To je průhledné a korektní pro věty, které nejsou komplikovány koordinacemi, elipsami a některými okrajovými jevy. **Závislostní strom** je struktura, která je konečným stromem ve smyslu teorie grafů, má kořen, do kterého směřují všechny cesty a jeho uzly jsou úplně (lineárně zleva doprava) uspořádány. Uzly (jistým, ne vždy stejným způsobem) reprezentují výskyty slov ve větě, hrany reprezentují vztah mezi rozvíjeným a rozvíjejícím slovem (jednotkou) ve větě.

Zbývá popsat, jak přecházet od redukční struktury k závislostnímu stromu. Redukční závislosti nepředstavují problém, příslušné hrany charakterizují vztah mezi slovem rozvíjejícím a slovem, které je rozvíjeno, pořadí slov ve větě je zachováno. U redukční komponenty musíme určit, které slovo z dané komponenty bude považováno za slovo řídicí a které slovo / která slova za závislá. K tomu je potřeba uvést další pravidla pro jednotlivé lingvistické jevy, které přiblížíme v následující sekci.

## 4 Redukční vztahy v přirozeném jazyce

Formální typologie závislostí zavedená v předchozí sekci odpovídá tradiční lingvistické klasifikaci – zde se pokusíme tuto souvislost detailněji popsat.

### 4.1 Krátce k lingvistickým pojmům

Přiblížíme ve stručnosti pojmy (české) lingvistiky, které budeme dále používat. Více lze nalézt například v [10], [9], [7] a [5].

<sup>7</sup> Za sousední slova zde považujeme vždy dvojici slov z jedné redukční komponenty, mezi kterými neleží slovo ze stejné komponenty (taková slova se nemusí vyskytovat na sousedních slovosledných pozicích ve větě, mezi nimi může být slovo, které do této komponenty nepatří).

**Stavba věty** zachycuje vztahy mezi jednotlivými větnými členy, a to zejména (i) vztah **podřízenosti** (anglicky **subordination**),<sup>8</sup> tj. vztah rozvíjený větný člen – větný člen jej rozvíjející, a (ii) **slovoslednou pozici** ve větě (viz [10]). Jako větné členy přitom označujeme výrazy (jednotlivá slova i skupiny slov), které mají jedinou syntaktickou funkci (tj. vyjadřují např. podmět, predikát, předmět, přívlástek či příslovečné určení).

Kritérium pro rozlišení větných členů ve dvojici **člen rozvíjený** – **člen jej rozvíjející** (anglicky **modified** – **modifying member**) je založeno na pojmu tzv. endocentrické konstrukce,<sup>9</sup> viz [9]: pokud lze jedno ze dvou slov vypustit, aniž se změní distribuční vlastnosti celého páru (tj. aniž se změní schopnost vyskytovat se ve stejném syntaktickém okolí), je toto slovo považováno za rozvíjející. Na základě **principu analogie** (na úrovni slovních druhů) se určí směr rozvíjení i pro tzv. exocentrické konstrukce<sup>10</sup> – ilustrujme takový postup v případě slovesa a jeho aktoru (což je větný člen v činné větě typicky odpovídající subjektu): protože existují slovesa, která nevyžadují aktor (např. *prší*), je vhodné považovat aktor vždy za větný člen rozvíjející sloveso. Obdobně je i ‚povinný‘ objekt (např. u slovesa *potkat*) považován za větný člen rozvíjející sloveso (neboť existují slovesa, která povinný objekt nemají, např. sloveso *zemřít*).

Jako **doplnění slovesa**, resp. **podstatného jména**, **přídavného jména** či **příslovce** budeme označovat větné členy, které rozvíjejí dané sloveso, resp. podstatné jméno, přídavné jméno či příslovec. Přitom rozlišujeme vnitřní a volná doplnění. **Vnitřní doplnění** (v anglické literatuře obvykle **inner participants** či **arguments**) odpovídají podmětu a předmětu/ům daného slovesa (např. *studenti dělali zkoušku, kursem prošli studenti, maminka pověděla dětem pohádku, král zemřel, Petr četl o neštěstí, rodiče čekali na děti*), resp. jistým typům přívlástků u podstatného jména (např. *začátek přednášky*) a jistým rozvitím u přídavných jmen a příslovcí (např. *závislý na počasí, kolmo na základnu*). **Volná doplnění** (anglicky obvykle **adjuncts**) odpovídají příslovečným určením (času, místa, způsobu, ..., např. *jde včas, jde domů, jde pomalu*), resp. jistým typům přívlátku (např. *malý stůl*), či jistým rozvitím přídavných jmen a příslovcí (např. *trochu krátký, velmi rychle, zcela jistě*).

Všechna doplnění dělíme na **obligatorní**, tj. povinně přítomná ve významové reprezentaci věty (která však nemusí být vyjádřena v konkrétní (povrchové) realizaci věty, posluchači / čtenáři mohou být známa z předchozího kontextu), a na

<sup>8</sup> Termínem ‚podřízenost‘ v tomto článku označujeme jazykový vztah, zatímco termín ‚závislost‘ vyhrazuje pro formální struktury, pomocí nichž modelujeme jazykové vztahy, tedy např. podřízenost. Obdobně např. Kunze v [4] rozlišuje ‚Unterordnung‘ pro jazykové vztahy a ‚Abhängigkeit‘ pro jejich formální zachycení.

<sup>9</sup> Endocentrická konstrukce je konstrukce sestávající alespoň ze dvou slov, z nichž jediné je ‚povinné‘ a tvoří ‚hlavu‘ konstrukce, ostatní slova jsou ‚nepovinná‘ a rozvíjejí tuto ‚hlavu‘, např. spojení *malý stůl* je endocentrická konstrukce, podstatné jméno *stůl* tvoří ‚hlavu‘, přídavné jméno *malý* je rozvíjící.

<sup>10</sup> Exocentrická konstrukce je konstrukce, která nemá ‚hlavu‘, která by mohla (syntakticky) zastupovat celou konstrukci (jde tedy např. o předložkové skupiny); z argumentace v [9] lze vyvodit, že za exocentrické konstrukce lze považovat i např. konstrukce sestávající ze slovesa a jeho ‚povinných‘ doplnění.

doplnění ‚nepovinná‘, **fakultativní**. Pro rozlišení obligatorních a fakultativních doplnění slouží **dialogový test** popsany v [7].

## 4.2 Modelování jazykových jevů pomocí redukční struktury

Vraťme se zpět k formální typologii redukčních vztahů a podívejme se, jakým způsobem souvisejí s různými typy vztahů jazykových.

**Redukční závislosti** dovolují přímo modelovat fakultativní volná doplnění – jde vesměs o endocentrické konstrukce, u nichž lze celou dvojici nahradit rozvíjeným slovem, ‚hlavou‘ konstrukce (a to beze ztráty významové samostatnosti, princip (d) DAR). Takto jsou tedy zachyceny vztahy typu *malý stůl, jde pomalu, jde domů, jde včas, trochu krátký, velmi rychle, zcela jistě*.

Slovo (v redukci) řídicí odpovídá rozvíjenému slovu ve větě (tj. ‚hlavě‘ endocentrické konstrukce), slovo v redukci závislé odpovídá slovu, které toto řídicí slovo rozvíjí (viz obrázek 7).

Zbývá určit řídicí a závislý člen zpracovávané redukční závislosti v případech, kdy je rozvíjený nebo rozvíjející člen této závislosti tvořen celou redukční komponentou.

(i) Pokud je rozvíjející člen tvořen redukční komponentou, potom jako závislý člen určíme řídicí slovo této komponenty (určení řídicího slova u redukční komponenty viz níže, zbylé členy komponenty budou tvořit podstrom s kořenem v příslušném řídicím slově).

(ii) Pokud je rozvíjený člen tvořen redukční komponentou, potom v obecném případě dochází k (významovým) nejednoznačnostem, které jsou pro češtinu dosti typické (zajímavé příklady jsou uvedeny v [2]).



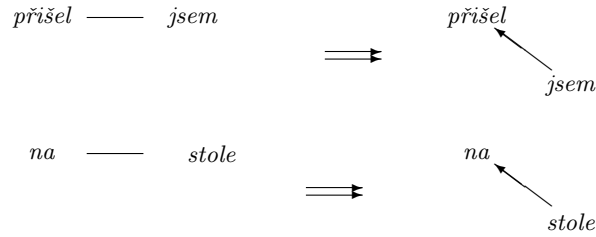
**Obrázek 7.** Redukční závislosti modelují volná doplnění.

**Redukční komponenty** dovolují modelovat složitější vztahy mezi výskyty slov. Jsou to jednak

- (a) vztahy morfologicko-syntaktické, jednak
- (b) vztahy syntakticko-sémantické.

**ad (a)** Redukční komponenty popisují tzv. **formémy**, což jsou jednotky odpovídající jednomu větnému členu – jsou to zejména předložkové skupiny (*před usnutím, na stole, vzhledem k Pavlovi*) nebo složené slovesné tvary (*přišel jsem, bude obědovat, je vytištěn, tiskne se*). (Jde tedy o exocentrické konstrukce.)





**Obrázek 8.** Možný převod formémů na závislostní podstrom (podle úzu PDT).

Formémy se v tradiční lingvistice znázorňují jako jeden uzel diagramu, resp. závislostního stromu znázorňujícího větnou stavbu, viz např. [10], resp. [9].<sup>11</sup> Pro mnohé prakticky orientované úkoly (např. kontrola gramatické správnosti, budování syntakticky anotovaného korpusu) je ovšem vhodné reprezentovat každé slovo věty vlastním uzlem (tedy nejen slova plnovýznamová); aby byl zachován zavedený datový typ závislostního stromu, je potřeba určit dodatečná pravidla, na jejichž základě se i redukční komponenty převedou na podstromy, tj. je nutné určit, které slovo formému se bude určovat jako řídicí a které/á jako druhotně závislé/á (vybrané řídicí slovo budeme pro jednoduchost označovat jako ‚hlava‘, stejně jako u endocentrických konstrukcí). Taková pravidla jsou obvykle technického charakteru a mohou se v jednotlivých projektech lišit (na obrázku 8 je řešení přijaté v Pražském závislostním korpusu, PDT).

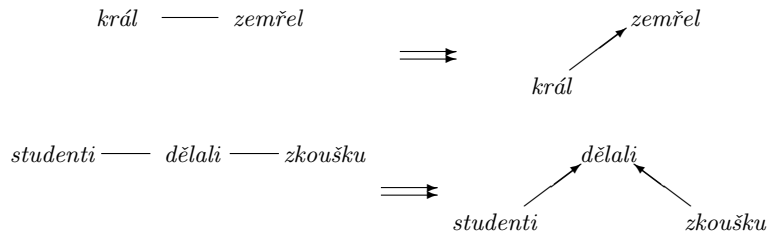
**ad (b)** Druhým typem vztahů, které jsou modelovány redukčními komponentami, jsou syntakticko-sémantické vztahy. Jsou to zejména **valenční vztahy** – vztahy slovesa, resp. podstatného jména, přídavného jména či příslovce a jeho povinných valenčních doplňků. Jde tedy o konstrukce typu *studenti dělali zkoušku*, *kursem prošli studenti*, *maminka pověděla dětem pohádku*, *král zemřel*, *Petr četl (o neštěstí)*, *rodiče čekali (na děti)*, *(na přednášku) přišli studenti*, *začátek přednášky*, *závislý (na počasí)*, *kolmo (na základnu)*<sup>12</sup>. Tyto konstrukce (bývají charakterizovány jako exocentrické konstrukce) nelze beze ztráty významové samostatnosti, princip DAR (d), nahradit jedním slovem, ‚hlavou‘ konstrukce.

Valenční vztahy tradiční lingvistika zachycuje pomocí závislostního stromu (viz [9], kde se vychází z [10]). Teoretickým kritériem pro určení rozvíjeného a rozvíjejícího větného členu se stal princip analogie, který zde byl krátce popsán v předchozích odstavcích. Na základě tohoto kritéria stanovujeme také pravidla pro určení řídicího slova při převodu redukční struktury na závislostní strom: sloveso určujeme jako řídicí větný člen, slovesná doplnění jako jeho slova závislá; obdobně pro podstatná jména, přídavná jména a příslovce a jejich doplňky.

Poznamenejme, že princip analogie zde lze jednoduše nahradit uvolněním požadavku (d) na zachování významové samostatnosti během DAR.

<sup>11</sup> V těchto pojetích jsou reprezentována samostatným uzlem pouze slova plnovýznamová (zejména významová slovesa, podstatná jména, přídavná jména a příslovce).

<sup>12</sup> Závorkou zde pracovně vyznačujeme formémy, viz bod (a).



Obrázek 9. Převod valenčních vztahů na závislostní podstrom.

## 5 Závěrečné poznámky

DAR dovoluje formulovat vztah základních syntaktických jevů: závislosti a slovosledu. To je nepostradatelné zejména pro modelování skladby jazyků s bohatou flexí a volným slovosledem, kde závislost a slovosled souvisejí velmi volně a u jednotlivých národních jazyků dosti odlišně (srovnejme s angličtinou, kde závislosti určuje (hlavně) velmi striktní slovosled).

Ukázali jsme, že závislosti odvozujeme ze dvou různých, nepřekrývajících se, jednoduše pozorovatelných, jazykově nezávislých jevů: z redukčních závislostí a redukčních komponent. Poukázali jsme na to, že lingvistická taxonomie jazykových jevů (česká) tomuto rozkladu (rozdělení) odpovídá. Přiblížili jsme tak lingvistický a inženýrský (množinový) pohled na danou problematiku.

## Reference

- Hajič, J. 1998. Building a Syntactically Annotated Corpus: The Prague Dependency Treebank. In: Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová (ed. E. Hajičová), Karolinum, CU Press, Prague, pp. 106-132.
- Holan, T., Kuboň, V., Oliva, K., Plátek, M. 2000. On Complexity of Word Order. In: Les grammaires de dépendance - Traitement automatique des langues (TAL), Vol. 41, No. 1 (q.ed. S. Kahane), pp. 273-300.
- Jančar, P., Mráz, F., Plátek, M., Vogel, J. 1999. On Monotonic Automata with a Restart Operation. Journal of Automata, Languages and Combinatorics, Vol. 4, No. 4, pp. 287-311.
- Kunze, J. 1975. Abhängigkeitsgrammatik. Volume XII of Studia Grammatica, Akademie Verlag, Berlin.
- Lopatková, M. 2003. Valency in the Prague Dependency Treebank: Building the Valency Lexicon. In: PBML 79-80, pp. 37-59.
- Lopatková, M., Žabokrtský, Z., Skwarska, K., Benešová, V. 2001. Tektogramaticky anotovaný valenční slovník českých sloves. UFAL/CKL, TR-2002-15.
- Panevová, J. 1980. Formy a funkce ve stavbě české věty. Academia, Praha.
- Plátek, M., Lopatková, M., Oliva, K. 2003. Restarting Automata: Motivations and Applications. In: Proceedings of the workshop "Petrinetze" (ed. Holzer, M.), Technische Universität München, pp. 90-96.
- Sgall, P., Hajičová, E., Panevová, J. 1986. The Meaning of the Sentence in Its Semantic and Pragmatic Aspects (ed. by J. Mey), Prague:Academia.
- Šmilauer, V. 1958. Učebnice větného rozboru. Skripta FF UK, SPN, Praha.