

Kapitola 1

Státnice I3: Závislostní syntax

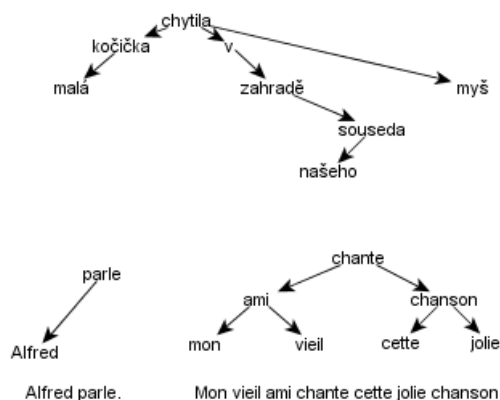
1.1 Úvod

Závislostní syntax je způsob popisu větné struktury, formálně zpracovaný L. Tesnièreem (dílo *Eléments de syntaxe structurale* vydáno 1959 posmrtně) v rámci tradice evropské strukturální lingvistiky (Tesnière spolupracoval i Pražským lingvistickým kroužkem). Na češtinu byla aplikována v knize *Novočeské skladba* (1947) V. Šmilauera a dále rozvíjena ve většině novějších mluvnic češtiny i v rámci *Funkčního generativního popisu* (FGD) P. Sgalla a dalších. Závislostní syntax se uplatňuje i v ruské teorii Meaning-Text.

Základní myšlenkou je vztah závislosti mezi jednotlivými slovy, přičemž v centru stojí hlavní sloveso a na něm závisí všechny ostatní členy (valence). Subjekt už tedy nemá stejné výsadní postavení jako v tradiční školské gramatice. Závislosti mezi jednotlivými slovy pak větu přirozeně uspořádají do stromové struktury.

Závislostní popis je vhodný i pro jazyky s volným slovosledem, na rozdíl od popisu pomocí bezprostředních složek, který na slovosledu do značné míry závisí.

1.2 Závislostní strom



Obrázek 1.1: Závislostní stromy uspořádané podle slovosledného pořadí (nahore) nebo podle závislostí (dole)

Závislostní strom se formálně definuje jako pětice $T = \langle N, Q, E, WO, L \rangle$, kde:

- (N, E) je orientovaný graf (strom, tj. souvislý a bez kružnic, každý uzel kromě kořene má právě jednoho otce),
- Q je množina možných ohodnocení uzlů (gramatické kategorie),
- $WO \subset N \times N$ je silné úplné uspořádání (určující pořadí slov)
- $L : N \rightarrow Q$ je ohodnocovací funkce)

Takto vytvořený závislostní strom má přesně tolik vrcholů, kolik je slov ve větě (na rozdíl od složkového stromu, kde slova ve větě představují jen listy). Nedává ale žádnou informaci o tom, jak byla věta vytvořena – zaměřuje se čistě na vztahy mezi jednotlivými členy.

Topologické (lineární) uspořádání uzlů může být řešeno tak, jak naznačeno ve formální definici, ale může být provedeno jen čistě na základě struktury, jak to bylo v původním Tesnièreově popisu (viz obrázek).

1.3 Vztahy v závislostní syntaxi

Mezi slovy ve větě mohou existovat dva základní vztahy:

- Závislost (determinace) – jedno slovo nějakým způsobem (významově) určuje druhé.
- Slova jsou na stejné úrovni, pak se jedná o koordinaci (několikanásobný větný člen), apozici (přístavek) nebo parentezi (vsuvku).

Základní závislostní strom zachycuje pouze první z nich, pro ostatní je třeba vytvořit speciální formalismus.

Závislost se typicky vyjadřuje hypotaxí (podřadností) a koordinační a podobné vztahy parataxí (souřadností). Existují ale i příklady, kde je to naopak:

Nechoď ven, nastyďneš. (determinace paratakticky), Otec s matkou šli (koordinace hypotakticky)

Závislost

Vztahy závislosti mezi slovy ve větě se formálně vyjadřují:

- kongruencí – shodou gramatických kategorií
- rekci – určení gramatických kategorií nadřazeným členem (např. sloveso vyžaduje urč. pád svých doplňení)
- juxtapozicí – přimykáním (tj. závislý člen se prostě nachází poblíž nadřazeného ve slovosledu)

Závislosti mezi slovy ve větě je možné analyzovat na základě principu redukce: postupně se snažíme větu redukovat a vypouštíme slova, jejichž vynecháním zachováme gramatickou korektnost věty. Slova, která mohou být vypuštěna v libovolném pořadí, na sobě nezávisí. Rozlišujeme dva druhy závislostí:

- endocentrická – je jednoznačné, co závisí na čem, pořadí možného vypouštění je jasně dáno
- exocentrická – nelze jednoznačně určit, co závisí na čem (např. v rámci předložkových skupin, doplňení vyžadovaných slovesem)

U exocentrických závislostí se řídící a závislý člen určují na základě analogií s jinými případy (tj. sloveso je řídící člen, i když bez některých doplňení nemůže stát samostatně). Funkci a nutnost výskytu jednotlivých závislostí blíže popisuje teorie valence.

Některé závislosti se ve větě nedají analyzovat jednoznačně, věta má pak více interpretací. Jde zejména o tyto situace:

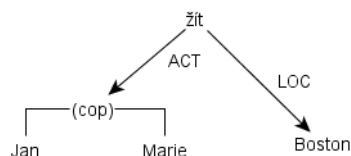
- tzv. “PP-attachment” (platí i pro složkové stromy, kde existuje více stromů odpovídajících stejné větě) a nejednoznačná doplňení adjektivem:

Př. Ředitel banky roku

- vyjádření užšího nebo volnějšího vztahu v některých konstrukcích (což nelze závislostním stromem, na rozdíl od složkového, popsat):

Př. Profesor zjistil, že je jeho (hladový algoritmus) nefunkční, (Zitřejší noviny) ze včerejška

Koordinace



Obrázek 1.2: Formalizace koordinace

Sémantický vztah koordinace (přiřadování) je druh “zmnožení”, obsazení jedné větné pozice více členy. Členy v tomto případě označují různé entity (větné členy nebo věty, i spojení větného členu s větou), které zastávají stejnou sémantickou roli. Jsou rovnocenné a samostatné, mají stejnou syntaktickou platnost. Do závislostí jdou jako celek, mají stejný řídící člen.

Podřadné a souřadné spojky se rozlišují podle pozice a přízvuku:

Př. **neboť** a **protože** se svojí souřadností / podřadností liší podle různých, spíš historicky daných kritérií (syntakticky jediný rozdíl: **neboť** nestojí nikdy na zač. souvětí).

Mezi členy koordinace nastává některý z následujících vztahů:

- kopulativní (CONJ – a)
- adverzativní (ADVS – ale)
- disjunktivní (DISJ – nebo)
- gradační: stoupá důležitost (GRAD – nejen, ale i) – v některých jiných lingv. tradicích se nerozlišuje
- příčiný (REAS – neboť), důsledkový (CSQ – a tak)
- oprava (spíše, lépe), zahrnutí (a to i)

Formalizace koordinace je složitá, je nutné přidat do závislostního stromu “další dimenzi”, nebo vkládat mezi závislostní hrany hrany “složkové”. Petkevič (1995) navrhl pro FGD použití dvou různých typů stromových hran pro každý ze vztahů, popř. v lineárním zápisu dva typy závorek. Všechny hrany jedné koordinace jsou navázány na jeden závislostní uzel (viz obrázek).

Apozice

Apozice je také zmnožení, kdy více větných členů má tutéž syntaktickou platnost. V tomto případě ale všechny pojmenovávají jeden referent. Jsou navzájem zaměnitelné a gramaticky kongruentní. Na existenci apozice se shodne většina popisů, jsou ale různá pojetí – např. Šmilauer považuje za apozici i výrazy **Pan Novák, Prezident Klaus**, kdežto v PDT je to přívlástek. Někdy se rozlišuje i zda je výraz oddělený čárkou, nebo uvedený v závorkách:

Př. "Obč. dem. strana (ODS) ...".

Apozici je taky občas problém odlišit od koordinace:

Př. "Naši sousedé, Marie a Milan, ..."

Formálně je možné ji zachytit stejně jako koordinaci.

Parenteze

Parenteze je vsuvka – věta nebo větný člen, jenž syntakticky nesouvisí s okolím, ale snaží se upřesnit, o čem se v okolní větě mluví. Typicky se zapisuje v závorkách, případně oddělená čárkou:

Př. Mohl byste, prosím, přijít?

Některé výrazy, jako např. **prosím**, **řekl bych**, se považují za ustálenou parentezi – de facto jde o částice nebo frazémy.

1.4 Projektivita

Uvažujeme závislostní stromy s uspořádanými vrcholy (např. podle slovosledu). Strom nad danou větou je projektivní, pokud neobsahuje žádnou neprojektivní závislost, tj. závislost mezi dvěma slovy oddělenými ve větě třetím slovem, které (ani nepřímou) nezávisí na žádném z nich. Pokud strom takovou závislost obsahuje, nazývá se neprojektivní.

Definice

Formálně definujeme pokrytí uzlu v závislostním stromě $Cov(u)$, $u \in T$ jako množinu všech indexů vrcholů (na základě úplného uspořádání), do kterých z uzlu u vede (orientovaná) cesta, tedy jsou na uzlu u přímo či nepřímou závislé. Do této množiny se počítá i samotný uzel u . Pro kořen platí $Cov(r) = \{1, \dots, |N|\}$.

Řekneme, že pokrytí uzlu $Cov(u) = \{i_1, \dots, i_k\}; i_1 < \dots < i_k; i_1, \dots, i_k \in \{1, \dots, |N|\}$ obsahuje díru, pokud existuje dvojice indexů vrcholů $(i_j, i_{j+1}) \in Cov(u), j \in \{1, \dots, k-1\}$ taková, že $i_{j+1} - i_j > 1$ (tj. pokrytí uzlu není souvislá řada indexů).

Potom pokud ve stromě existuje uzel, jehož pokrytí obsahuje díru, nazývá se strom neprojektivní.

Ekvivalentní definice projektivity říká, že jsou-li dva uzly u a v spojeny hranou a u leží nalevo od v (tj. má nižší pořadí ve slovosledném uspořádání), pak všechny uzly ležící nalevo od u a napravo od v jsou spojeny s kořenem cestou, která prochází jedním z vrcholů u, v .

Vlastnosti

Neprojektivní věty (tj. věty, jejichž syntaktickou analýzou je neprojektivní strom) jsou v některých jazycích spíše výjimkou (např. v angličtině), v češtině jsou naprosto běžné:

Př. *Karla jsme chtěli poslat do Francie. Soubor se nepodařilo otevřít.*

Př. *I saw a man with a dog yesterday which was a yorkshire terrier.*

Neprojektivní konstrukce nelze zobrazit souvislým složkovým stromem. V závislostním stromě to možné je. Ve FGD se s neprojektivitou počítá na úrovni povrchové syntaxe, ale už ne na tektogramatické rovině. Pořadí uzlů v tektogramatickém stromě totiž neodpovídá pořadí podle slovosledu.

1.5 Valence

Valence je vlastnost lexikálních jednotek (slov), která drží mnoho závislostních vztahů pohromadě. Jedná se o schopnost slov (prototypicky sloves, ale v mnoha teoriích se valence přiznává i jiným slovním druhům) vázat na sebe jiná slova, čímž vzniká větná struktura. Popisoval ji už Tesnière a přijímají ji v podstatě všechny závislostní teorie. Souvisí s principem redukce – jednotky, jejichž pozice se díky danému slovu otevírají, považujeme za závislé.

Valence se snaží popsat situaci, která se nám vybaví v souvislosti s daným slovem – zachytit, které sémantické participanty mají být přítomny (termín sémantický participant se nedá dobře definovat, ale všichni si v praxi představí to samé), tedy popsat počet a povahu argumentů, které na sebe slovo váže.

Valence pomáhá rozlišit nejednoznačnosti v:

- morfologii: *Ptala se jeho bratra* (Gen. nebo Acc.?)
- syntaxi: *Začala ho milovat. Nechala ho spát* (na čem závisí ho?)
- významech slova: *odpovídat na / za / čemu*
- sémantice větných členů: *sháněl se po ... / přišel po ...*

Je důležitá pro zpracování přirozeného jazyka, hlavně pravidlovými metodami (ve statistických je přítomna implicitně), učení se jazyku a lingvistický výzkum.

Pojetí valence se v jednotlivých teoriích liší. Jde už o rovinu lingvistického popisu, na které se uvažuje – pro FGD je to tektogramatická rovina, ale jiné teorie (německé, Meaning-Text) zahrnují valenci i do povrchové syntaxe. Týká se to i druhu vztahů, které jsou do valence zahrnuty – ve FGD valence souhlasí s hranami závislostních stromů, teorie Meaning-Text chápe jako valenci i např. to, že adjektivum vyžaduje substantivum, které by rozvíjelo (“pasivní valence”).

Popisy valenčních rolí, tj. ohodnocení jednotlivých valenčních závislostí, se taky značně liší napříč teoriemi. Některé používají velice jemné sémantické rozdělení (FrameNet), jiné hrubší (německé teorie, Case Grammar a podobně i θ -role v Chomského teorii), jiné používají kombinace syntaktické a sémantické klasifikace (FGD, Meaning-Text). Některé teorie, jako např. původní Tesnièreova syntax, nebo z části PropBank/NomBank, dokonce doplnění pouze číslují.

Mnoho popisů valence se ale shodne na základním rozdělení dvou druhů závislostí, a to:

- Aktanty (valenční argumenty, vnitřní doplnění, participanty) – role, které mohou (nebo musí) být pro určitý řídicí prvek obsazeny pouze jednou, navíc je jejich množina pro danou lexikální jednotku typická, tj. dají se vyjmenovat.
- Volná doplnění (adverbiální modifikátory, circumstantials) – role, které se mohou opakovat i několikrát a navíc určité doplnění se může vyskytovat s libovolným řídicím slovem.

Např. v Case Grammar se ale takovéto rozdělení neprovádí.

Typicky se taky u jednotlivých doplnění rozhoduje o jejich:

- sémantické obligatornosti nebo fakultativnosti – nějaké doplnění musí být buď vyjádřeno, nebo známo z kontextu (pozná se dialogovým testem)
- syntaktické vypustitelnosti nebo nevypustitelnosti – je-li nějaké nevypustitelné doplnění nevyjádřené, věta není gramaticky korektní

Typicky jsou aktanty obligatorní a volná doplnění fakultativní, ale není to jediný možný případ:

Př. *Jan se chová hezky.* (obligatorní, nevypustitelné volné doplnění), *vyrábět něco* (z něčeho) (fakultativní aktant)

Tato konkrétní terminologie pochází z FGD, ale podobné koncepty se vyskytují i jinde.

Valenci slov nelze zachytit pravidly, uchovává se ve slovníku. Takový slovník už zachycuje hodně syntaktických informací.