# Kapitola 18

# Státnice I3: Funkční generativní popis

## 18.1 Úvod

Teorii začal vyvíjet P. Sgall na FF UK na zač. 60. let, inspirovaný Chomského teorií a motivovaný strojovým překladem. Vycházel ale přitom z tradic Pražského lingvistického kroužku a strukturalismu – centrem je tedy jazykový systém (langue), klade se důraz na explicitní formalizaci a celé je to založeno na syntaxi.

### Zákl. koncepce

Základní rysy teorie jsou:

- Závislostní přístup, valence (sloves a i dalších slovních druhů) (J. Panevová)
- Stratifikace (rozložení popisu na jednotlivé roviny podle úrovně abstrakce)
- Vztah formy a funkce (jedna forma má více funkcí na vyšších rovinách, jedna funkce více forem na nižších (asymetrický dualismus))
- Jazykový význam (vyloučení kognitivního obsahu z popisu; rozlišení víceznačnosti a zachování vágnosti)
- Aktuální členění jako součást významu (P. Sgall, E. Hajičová)

### Generování

Původní koncepce vychází z představy, že generování (tvorba vět na základě významového popisu) bude jednodušší než analýza. V původní verzi bylo generování rozdělené na dvě hlavní fáze:

- Generativní složka – vymezovala (v původní verzi pomocí přepisovacích pravidel frázové gramatiky a složkových stromů, které ale měly indikaci směru a typu závislostí) správné zápisy vět na tektogramatické rovině
- Překladové složky – přepis na nižší úrovně až do běžného textu (formálně 4 zásobníkové automaty a jeden regulární)

Kvůli frázovým stromům se generoval jen jeden druh slovosledu apod. V pozdějších verzích byly frázové stromy upraveny na závislostní. Skutečně existovala v 70. – 80. letech implementace, která generovala korektní české věty, ale dodnes se nedochovala.

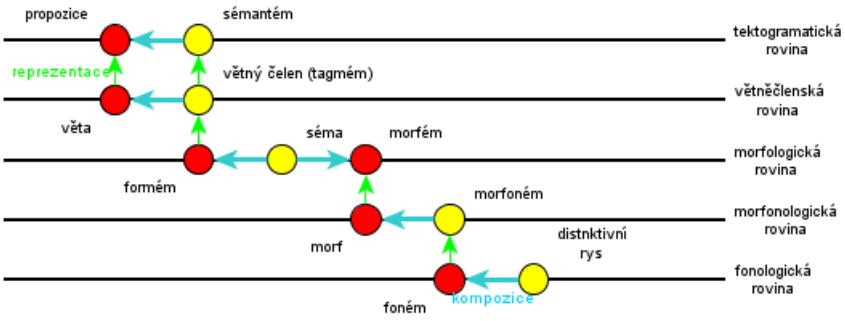
## 18.2 Roviny popisu

Popis jazyka se tu uskutečňuje na několika rovinách abstrakce, od lineárního proudu hlásek po samotný význam. Na každé rovině je ale reprezentovaná celá věta – v nižších lineární strukturou, na vyšších závislostním stromem.

- forma, funkce – nižší rovina je formou vyšší roviny (vztah reprezentace), základní jednotky na jedné rovině tvoří komplexní (kompozice)

Teorie FGD obsahuje tyto roviny popisu (od nejvyšší k nejnižší, hlavně v nižších úrovních není počet ustálený):

- tektogramatická (hloubková syntax, rovina jazykovéhyo významu)
- povrchová syntax (od 90. let Sgall zpochybnil její nutnost, v komputační lingvistice se z praktických důvodů stále používá)
- morfematická (morfologická)




Obrázek 18.1: Roviny popisu FGD: základní jednotky rovin jsou značeny žlutě a složené červeně

- morfonologická
- fonologická / fonetická

Platí, že nižší rovina je formou vyšší roviny a vyšší rovina funkcí nižší (vztah reprezentace). Na každé rovině existují základní jednotky popisu, které dávají dohromady složitější (vztah kompozice). Složitější jednotky pak zpravidla slouží jako základní na vyšší rovině.

Tektogramatická rovina musí obsahovat všechnu významovou informaci, během převodu na nižší roviny se žádná už nedodává. Základní jednotkou (uzly stromu) jsou sémantémy, celek se nazývá propozice. Ohodnocení sémantémů sestává z komplexního symbolu, lineární řazení jde podle "hloubkového slovosledu" – dynamiky výpovědi (aktuálního členění). Komplexní symbol sestává z následujících informací:

- lexikální informace – měla by obsahovat ne povrchový lexém, ale tektogramatický – synonyma by měla být ztotožněná, slovesná podstatná jména zahrnuta pod slovesa atd. (ale v PDT to tak úplně není)
- morfologická informace – jde taky o význam: mluvím o jednom, nebo více objektech? kdy se odehrává děj? (jen když si mluvčí vybírá, např. kongruence nás nezajímá)
- syntaktická informace – pomocí funktoru vyjadřuje vztah rodiče a dítěte ve stromu (ACT, PAT ... atd.)

Větněčlenská rovina pracuje s tagmémy (větnými členy), jejich kompozicí vzniká věta. Na morfologické rovině se z jednotlivých sémat (Sg., Nom., Fem. apod.) skládají jednak morfémy, odpovídající morfům na morfonologické rovině a jednak formémy (např. slova, předložkové vazby apod.), které odpovídají tagmémům. Morfémy se dělí na lexikální (kmeny, odvozovací předpony a přípony) a gramatické (ty vyjadřují zpravidla více sémat).

Fonému z fonetické roviny odpovídá morfoném (tj. všechny alofony v daném místě daného morfému). Kromě řetězů morfonémů – morfů – obsahuje morfonologická rovina i nástroje pro zachycení suprasegmentálních jevů (přízvukový takt, věta – intonace). Fonémy se skládají na fonologické/fonetické rovině z distinktivních rysů. Fonetická rovina se často z popisu vynechává (někdy zas se naopak ponechává a vynechává se morfonologická), lze taky nahradit fonologickou a fonetickou rovinu rovinou grafématickou.

## 18.3 Jazykový význam

Pro popis na tektogramatické rovině ve FGD se ostře odlišuje jazykový význam od myšlenkového obsahu (kognitivního obsahu, primárně nejazykového), tj. popisujeme jen to, co je obsaženo v jazyce – strukturu specifickou pro daný jazyk, včetně pragmatických rysů (indexy), ale zbavenou synonymie, homonymie a dalších nepravidelností. Už Saussure označoval význam za "formu obsahu". Rozlišuje se víceznačnost, naopak zachovává se vágnost.

Význam je neformálně to, co je viditelné přímo z formy vyjádření, obsah už jsou vyvozované výroky (v praxi je to často horší odlišit).

I pro rozlišení víceznačností je někdy třeba věcných znalostí:

Př. `Chytil tlouště na višni.` – musíme vědět, že neseděl na višni, ale že jde o návnadu

Pořád se ale jedná o víceznačnost, protože jde ale o jazykový fenomén (homonymie dvou různých doplnění).

Význam je vázaný na syntaktické i lexikální elementy:

Př. `wash = mýt / prát`, `go = jít / jet` – v angličtině to skutečně je jeden význam toho slova, není tam dvojznačné

Př. `fingers / toes = prsty` – totéž v češtině (prsty jsou to všechny, musí být blíže specifikovány rozvitím nebo kontextem)

Totéž platí např. o kategorii vidu, která se nekryje přesně s jinými vyjádřeními (časy v angličtině, lexikální prostředky v němčině apod). Jiné podobné fenomény jsou např. odlišení duálu nebo rozlišení osob "my včetně tebe" a "my kromě tebe" v některých jazycích.

Vágnost je naopak vlastní významovým jednotkám každého jazyka, vlastností významu je být vágní. Její rozlišení už není předmětem jazyka (a tedy popisu ve FGD), ale myšlenkového obsahu:

Př. `Francouzi nejedí polévku.` – že jde o "typické Francouze", věta neudává

Př. `Děti dostaly dárky.` – neříká se, kolik dárků dostalo které dítě

Vágní jsou i relační adjektiva – `švestkové / bramborové knedlíky` – nebo přechodníky. Vágnost je i v časové souslednosti vět v češtině:

Př. `Od té doby, co matka zemřela, bylo nám stále hůř.` – neříká se, jaký je vztah dvou vět, ale dá se pochopit, že následný

Většina vágních konstrukcí lze pochopit z kontextu nebo "vyrozumět" vyvozováním důsledků.

Z podobných důvodů, jako se omezujeme na význam, se u určování funktorů ACT a PAT omezujeme na syntaktické kritérium (viz dále) – často je jejich detailní sémantika totiž vágní a lze pouze "vyrozumět" z okolí.

Př. `Otec otevřel dveře. Klíč otevřel dveře. Vítr otevřel dveře.` – toto můžeme považovat za vágnost

## 18.4 Valence

Ve FGD je valence zkoumána už od počátku. Úzce se týká významu slov, proto se řadí na tektogramatickou rovinu. Dotýká se ale i nižších vrstev, protože valenční doplnění mohou vyžadovat konkrétní formu.

Každý autosémantický slovní druh je charakterizován valencí (frame-bearing words), primárně se jedná o slovesa, ale valenci lze nalézt i u substantiv, adjektiv, a adverbií.

Př: `zájem o co, bratr koho, předělaný z čeho na co, kolmý na co, blízko čeho`

Pro slovesa je ovšem teorie nejpropracovanější, nejpřesnější. V jiných teoriích se mluví i o valenci předložek, ale ve FGD to nemáme – to, že předložka dává pád substantivu, považujeme za morfologický jev (rekci).

### Doplnění

Ve FGD se valenční doplnění dělí na obligatorní a fakultativní – obligatorní musí být (na tektogramatickérovině) vždy přítomna, abychom měli sémanticky úplný a srozumitelný zápis (nemusí být ale vyjádřena). Jejich přítomnost se prokazuje dialogovým testem:

Př. `A: Moji přátelé přijeli. B: Odkud? A: Nevím.` – OK, `B: Kam? A:*Nevím.` – nelze, proto určení kam je obligatorní

Některá doplnění jsou syntakticky nevypustitelná (tj. musí být vždy vyjádřená), jiná jsou vypustitelná.

Dále se doplnění dělí na aktanty a volná doplnění. Ve FGD se do valenčního rámce nějakého slova zapisují všechny aktanty (vč. fakultativních) a obligatorní volná doplnění (např. pro slovesa `přijít, chovat se`).

#### Pojetí aktantů u sloves ve FGD

Ve FGD Máme 5 aktantů, definovaných spíše syntakticky – ACT a PAT téměř výhradně, ostatní (EFF, ORIG, ADDR) částečně sémanticky. Kvůli svému spíše syntaktickému určení mají ACT a PAT hodně sémantických funkcí.

Jde o kompromis mezi hodně sémantickým přístupem, jako má např. FrameNet C. Fillmorea (doplnění jsou dnes pro každou typizovanou skupinu sloves jiná, hodně detailní), a hodně syntaktickým, jako obsahuje PropBank (aktanty jsou číslované a číslování specifické pro každé slovo). Syntaktický přístup k valenci prosazoval už Tesnière, z něj FGD vychází – např. akademická mluvnice češtiny (Daneš) razí naproti tomu sémantický přístup.

U aktantů se ve FGD uplatňuje princip posouvání:

- první aktant je vždy ACT
- druhý vždy PAT
- třetí je ADDR, ORIG nebo EFF
  - když nelze rozhodnout sémanticky, je to EFF

Př.: `Petr(ACT) vyrostl z chlapce(ORIG) v mladého muže(PAT!)`

Př.: `The janitor(ACT) opened the door(PAT) with a key(MEANS). A key(ACT) opened the door(PAT). The door(ACT) opened.`

EFF má primární význam "výsledek děje", nebo "vlastnost přiřazovaná patiensu" (`vyprávěl o nich, že...`). ADDR a ORIG jsou sémanticky homogenní, skoro jako volná doplnění. ADDR představuje příjemce informace nebo předmětu (i odebrání). ORIG značí látku původu, původce předmětu nebo informace při výměně:

Př.: `Dům je z kamene(PAT!). Vyrobil něco z něčeho(ORIG). Dozvědět se něco(PAT) od někoho(ORIG)`

ADDR a ORIG se jen málokdy dají dobře zkombinovat, ale existují i slovesa, kde je všech pět aktantů možných:

Př.: `Maminka(ACT) předělala dětem(ADDR) loutku(PAT) z kašpárka(ORIG) na čerta(EFF).`

## Valenční informace ve slovníku

Valenční informace se uchovávají ve slovníku. Typicky patří valenční rámec k jednomu významu slova ((základní) lexikální jednotce), proto jeden lexém (soubor všech významů a forem příslušných jednomu základnímu tvaru – lemmatu) může obsahovat několik rámců.

Valenční slovník by se měl dělat z dat a ručně. Ukazuje se, že malý počet sloves pokryje velkou část korpusu, jen málo slov má větší počet lex. jednotek. Zároveň se různá slovesa chovají různě a mají různé rámce, i když popisují úplně stejnou sémantickou situaci; umožňují vyjádřit různé participanty.

V teorii FGD jsou zpracované slovníky PDT-VALLEX (pro Pražský závislostní korpus, jejž pokrývá) a VALLEX (ten se snaží o komplexní popis všech významů slov, slovesa jsou vybrána na základě frekvence v Českém národním korpusu).

## Valence substantiv a adjektiv

Všechna valenční doplnění substantiv a adjektiv bývají vypustitelná. Liší se podle toho, zda se jedná o primární nebo deverbativní substantiva nebo adjektiva.

### Primární substantiva

Rozlišují se následující doplnění:

- Partitiv/materiál (aktant) – množství/skupina (`dvojice, balení, sada`), kontejner (`sklenice, pytlík, tisíc`)
- Přináležitost (volné, u relačních substantiv (`otec, příbuzný, nadřízený`) aktant) – příbuzenský vztah, vztah části a celku (`střecha domu`), nositel vlastnosti (`míra čeho, délka čeho, čí upřímnost`), vlastnictví, přináležení (`klíč od`)
- Identita (volné) – metajazykové výrazy (`agentura Reuters, pojem času`), i další (`nápis Obětem války`)
- Autor (volné)
- Přívlastek restriktivní (volné)
- Přívlastek deskriptivní (volné)

### Deverbativní substantiva

Pro valenční chování je důležitý typ derivace, jakým vznikly:

- syntaktická derivace – čistě syntaktický prostředek: `dělání, pokrytí`. Je tu vidět původní valence, ale často dochází k abstrakci (nevyjádření některých původně povinných aktantů)
- lexikální derivace – vznik ze sloves (základové slovo), ale sémanticky jde skutečně o substantiva: `letec, letiště`. Substantivum samo vyjadřuje jeden z participant děje – toto doplnění mizí (zabudování pozice), ostatní jsou přípustná, ale často taktéž nevyjádřená, dochází k uvolnění vazeb, často některá doplnění ani vyjádřit nelze (zní to divně).

Nejde o vyhraněné dělení, spíše škálu, přechod – je i spousta případů "mezi" (`dar, let` – "široce dějová jména").

Při konverzích ze sloves dochází ke změnám morfologického vyjádření, strukturní pády (Nom., Acc.) se primárně mění na genitiv:

Př. `vyrábět něco -> výroba čeho`

To ukazuje, že možnost vyjadřovat doplnění je u substantiv omezenější (genitivy se typicky nesmí opakovat). Nestrukturní pády (zejména Dat., Ins., ale i Gen., předložkové pády, infinitiv) většinou zůstávají, adverbia se mění typicky na adjektiva. Nepravidelnosti ale existují:

Př. `zájem o něco / na něčem`, `strachovat se čeho -> strach z čeho`

#### Primární adjektiva

Mají stejný repertoár možných doplnění jako slovesa, navíc komparativ má `než` a superlativ `z koho/čeho`. V teorii se zde už nepočítá s posouváním, ADDR, PAT se rozlišuje sémanticky. Většina adjektiv má jen jedno doplnění, jen výjimky více (`nápadný čím komu, vděčný komu za co`). Prototypicky se nevyskytuje ACT.

#### Deverbativní adjektiva

Sloveso se mění na adjektivum, které rozvíjí jedno z původních valenčních doplnění. Zachovávají rámec sloves až na jeden aktant, který je obsazený rozvíjeným substantivem. Povrchově jsou všechna doplnění vypustitelná.

Př.: např. `[kdo] omezí [co na co] -> *kdo* omezený [kým na co]`

#### Adverbia

Mají valenční chování, ale nikdo ho zatím podrobně nestudoval.

Př: `kolmo na co, vedle čeho, blízko čeho`

## 18.5 Aktuální členění ve FGD

Aktuální členění je ve FGD popisováno už od první verze. Navazuje tak na tradici strukturalismu a Pražského lingvistického kroužku, zejména práce Viléma Mathesia a Jana Firbase.

### Definice

V různých teoriích najdeme různou terminologii, někdy se termíny kryjí přesně, někdy ne docela. Informační struktura věty je totéž co aktuální členění věty (původní termín od Mathesia), anglicky topic-focus articulation (podle ÚFALu, P. Sgalla a dalších), nebo functional sentence perspective (podle Brněnské školy, J. Firbase a dalších). Jde o dělení věty na:

- základ, východisko, téma věty nebo topic, tj. to, o čem se ve větě mluví (známá informace).
- jádro, ohnisko', réma nebo focus, tj. to, co se ve větě říká nového o známé informaci.

V pražském moderním přístupu se používá spíš anglických výrazů topic, focus a topic-focus articulation, protože původní české jsou zatíženy nepřesnostmi.

### Vyjádření aktuálního členění

Informační strukturu lze vyjádřit různými prostředky, v češtině hlavně slovosledem a intonací – intonace je velmi důležitá, i když máme volný slovosled (a intonace má i další funkce). V angličtině např. je intonace kvůli pevnému slovosledu ještě důležitější.

Př.: `John gave me a letter. I met him [in a bookshop] [yesterday].` – jestli je focus `yesterday` nebo `in a bookshop`, poznáme jen podle intonace.

Př.: `Nejdražší je Audi. / Audi je nejdražší.` – při normální intonaci je focus na konci, proto první věta odpovídá situaci, kdy mluvím o cenách vozů, kdežto druhá připadá hovoru o autech a jejich vlastnostech.

V češtině můžeme topic-focus articulation rozlišit např. i použitím krátkého nebo dlouhého tvaru zájmene (ve focusu budou spíše dlouhé tvary, dlouhé tvary zájmen se ale využívají i pro vyjádření kontrastu v rámci topicu).

Př.: `Dej mi tu knížku. / Tu knížku dej mně.`

V angličtině se dá informační struktura vyjádřit i použitím určitého nebo neurčitého členu.

Př.: `A disabled man limped inside. / The disabled man limped inside.` – v prvním případě je invalida ve focusu, v druhém v topicu

Můžeme použít ale i různé částice (focalizers, rematizátory, přitahující větný přízvuk) nebo speciální syntaktickou konstrukci, tzv. vytýkací (to bývá častější v angličtině).

Př.: `Teprve Jeník dokázal draka porazit.` – Jeník je díky částici `teprve` ve focusu.

Př.: `Bill introduced John only to SUE. / Bill introduced only JOHN to Sue. / Bill only INTRODUCED John to Sue. / Only BILL introduced John to Sue.` – rematizátor mění focus

Př.: `Byla to vichřice, co ho zničilo.` – vytýkací konstrukce, ve focusu je `vichřice`.

## Aktuální členění a význam

Aktuální členění úzce souvisí s funkcí sdělení, projevuje se ale různými formami (povrchovými strukturami věty); jedna forma může vyjadřovat naopak více různých aktuálních členění, ač to není tak časté:

Př.: `"Why do we dress boys in blue and girls in red?""Because they can't dress themselves."`

Aktuální členění patří do popisu významové stavby věty, v pražském popisu na tektogramatickou rovinu, protože jeho změna může změnit význam celé věty, když dojde ke změně presupozice – nutně předpokládané skutečnosti, aby měla věta smysl:

Př.: `The king of France didn't visit the exhibition. / The exhibition was not visited by the king of France.` – první varianta presuponuje existenci výstavy i krále, kdežto pro druhou nemusí francouzský král existovat.

Př.: `Aspoň dva jazyky zná v této místnosti každý. / Každý v této místnosti zná aspoň dva jazyky.` – První věta presuponuje dva stejné jazyky, ale druhá už ne.

Aktuálním členěním lze také měnit dosah negace (scope of negation) (negace může být buď v základu, nebo v jádře – potom se vztahuje na přísudek jen tehdy, je-li ten také v jádře):

Př. (1): `Moje sestra nehubovala bratra kvůli špatné známce` = `nehubovala vůbec / hubovala někoho jiného / hubovala bratra kvůli něčemu jinému.` – `moje sestra` nemůže být dotčeno negací, která je v jádru; stojí v základu

Př. (2): `Jirka nepřišel, protože mu došly peníze` – ve chvíli, kdy Jirka `nepřišel, ne` např. `protože byl nemocný, ale protože mu došly peníze`, se dostává negace do základu. Je to ale dvojznačné, můžu říct, že Jirka `přišel, protože chtěl vidět Marii` a potom je negace v jádru.

Různé druhy negace pak ovlivňují i presupozici:

Př.: `Jirka nezpůsobil naši porážku. / Naši porážku nezpůsobil Jirka.` – první věta je dvojznačná, kdežto v druhé je jasné, že jsme byli poraženi. Porážka se tak stává presupozicí.

Aktuální členění má navíc vliv i na alegaci věty (výrok, který vyplývá z kladné verze věty, ale ze záporné nevyplývá ani on, ani jeho negace). Může měnit presupozici v alegaci a naopak:

Př.: `Milanovou dceru včera viděl Jirkův bratr. / Včera Jirkův bratr viděl Milanovu dceru.` – v první větě se presuponuje existence Milanovy dcery a Jirkův bratr je jen alegován, kdežto v druhé větě tomu je přesně naopak.

Nejde jen o negace, ale i o kvantifikátory:

Př.: `Pražané většinou jezdí na Slapy. / Na Slapy jezdí většinou Pražané.` – v první větě neříkám, kdo všechno jezdí na Slapy, ale v druhé ano.

Pro nalezení změny ve významu při změně aktuálního členění nepotřebuju ale ani kvantifikátory:

Př.: `Na Moravě se mluví česky. Česky se mluví na Moravě.` – první případ je tzv. exhaustive listing – podávám úplnou informaci, protože na Moravě se jinak než česky nemluví; druhý ale ne, protože Česky se mluví i Čechách.

Př.: `Dogs must be CARRIED. / DOGS must be carried.` – první verze intonace říká, že mám-li psa, musím ho nést, druhá přikazuje nosit s sebou nějakého psa.

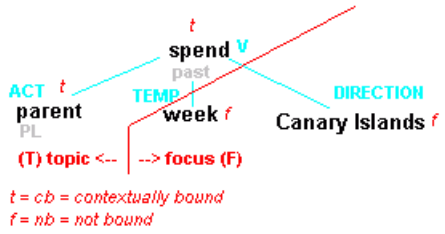
## Začlenění do FGD

Ve FGD se na tektogramatické rovině jednotlivé uzly závislostního stromu řadí podle dynamiky výpovědi – míry kontextové zapojenosti ("hloubkový slovosled"). Uzly si nesou informaci, zda jsou:

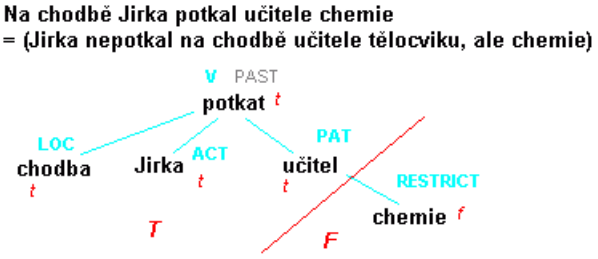
- kontextově zapojené (contextually bound) (značí malým písmenem t)
- nezapojené (not bound) (značí se malým f)

Na základě toho můžeme popsat, co je topic a co je focus.
Byla vypracovaná i pravidla pro oddělení topicu a focusu celé věty (značí se velkým T a F) podle těchto indikátorů:

1. Začne se od kořene (slovesa)
2. Přímé na slovese závislé složky vždy patří do T / F vcelku, se všemi svými členy dohromady (až na následující výjimku)




Obrázek 18.2: Ukázka topic-focus articulation podle funkčního generativního popisu




Obrázek 18.3: Pro oddělení topicu a focusu věty jsou nutná všechna 3 pravidla

3. Pokud jsou všechny přímé závislé složky kontextově zapojené, sleduje se podstrom poslední z nich (v pořadí členů ve větě), dokud se nenajde nezapojený element. Jeho podstrom je pak focus.

Mít jen první dvě pravidla nestačí (viz obrázek). Také to není jen výměna t a f za T a F, to platí jen na první závislé vrstvě, dá se to ukázat i na příkladu:

Př.: `Which schools do your children attend?` → `All (f) my children (t) attend (t) a private school (f) in London (f).` – v této větě je all sice kontextově nezapojené, ale patří do T.

V praxi byl tento algoritmus úspěšně zkoušen na větách z PDT.

### Kontrastivní zapojení a souvětí v PDT

Pro popis v PDT musela být teorie trochu rozšířena. Byly tam zahrnuty:

1. koordinace klauzí – každá koordinovaná klauze má vlastní aktuální členění.
2. subordinace – závislé klauze jsou součástí aktuálního členění hlavní klauze, ač mají i svoje vlastní podřízené aktuální členění. Stojí-li podřízená klauze v topicu, většinou jí souvětí začíná, stojí-li ve focusu, souvětí jí zpravidla končí.
3. kontrastivní zapojenost – kromě t a f se přidává c pro kontrastivně zapojené větné členy. c i t patří do topicu (T).

Př.: `Kde jsi se setkal se svými spolužáky?` → `Jirku (c) jsem viděl v divadle (koordinace klauzí), Andulu (c) na koncertě.`

## 18.6 Pražský závislostní korpus

Pražský závislostní korpus (ve verzi 2.0) obsahuje necelé 2 milióny tokenů (novinových článků z Českého národního korpusu) anotované na základě FGD, z toho asi 800 000 až do úrovně tektogramatické roviny. Z časových, finančních a implementačních důvodů je tam ale několik rozdílů oproti původní FGD. Hodně z nich pramení z obráceného pohledu (analýza místo generování).

Máme tři roviny popisu + jednu "nultou" (tj. čtyři):

- w-rovina (slovní, nultá)
- m-rovina (morfologická)
- a-rovina (analytická, původní "větně členská")
- t-rovina (tektogramatická)

Nultá rovina obsahuje jen jednotlivá slova (slovní tvary = tokeny – včetně interpunkce), tak jak šla za sebou v novinách. Není to tedy přesně text z novin, ač s původními překlepy, prošel už tokenizací – tedy ani vytvoření nulté roviny není triviální.

Morfologická rovina zhruba odpovídá FGD, tektogramatická taky, ale analytická původní větněčlenské moc neodpovídá. M-rovina a a-rovina si odpovídají slovo od slova. U nulté roviny a a m-roviny nemusí být vztah 1:1 vždy, kvůli překlepům (např. zapomenutá mezera), ale děje se to jen zřídka, bývá buď 1:m, nebo n:1, když už (teoreticky vyloučit m:n nelze).

Anotační formát je jazyk PML založený na XML (nadefinované datové typy apod.), zpracovatelný editorem TrEd. Popis jednotlivých rovin je oddělený a roviny jsou provázány odkazy.

## m-rovina

Sémata jsou sdružena ne do morfémů jako ve FGP, ale do slov. Text je rozdělen na jednotlivé věty; gramatické kategorie jsou reprezentovány tagy (morfolog. značkami). Překlepy jsou už opravené. Každý tvar má přiřazené lemma (slovníkový tvar), tam je taky rozlišená homonymie a přidané poznámky o vytvoření slova, např. "oživení" je z "oživit".

Morfologické tagy jsou poziční, s celkem 15 možnými údaji: slovní druh, jemněji určený slovní druh, rod, číslo, pád, rod vlastníka, číslo vlastníka, osoba, čas, stupeň, negace, slovesný rod + 3 rezervované (z nichž se používá poslední pro stylistické příznaky).

## a-rovina

A-rovina už se skládá ze stromů. Ty mají drobné odlišnosti, např. kořen je speciální, technický, na něm teprve závisí hlavní sloveso (kvůli nevětným konstrukcím, parentezím apod.).

Uzly jsou ohodnocené analytickou funkcí, která popisuje syntaktické kategorie (Pred, Pnom (přísudek jmenný), AuxV, Sb, Obj, Atr, Adv, AuxP (předložka), AuxC (podřadící spojka) atd.). Členové koordinací, apozicí a parentezí jsou speciálně označeni a v TrEdu je možné určit efektivního rodiče nebo efektivní děti každého uzlu.

## t-rovina

T-rovina docela dobře odpovídá tektogramatické rovině. Obsahuje tedy i uzly pro nevyjádřené větné členy, které jsou ve významu přítomné, neobsahuje uzly pro gramatická slova – jeden uzel tak může odkazovat k více uzlům analytické roviny. Uzlů je několik typů – pro běžná (vyjádřená) slova, pro nevyjádřená slova, koordinace a apozice, rematizátory apod.

Uzly mají určený funktor (tj. druh aktantu nebo volného doplnění, pro některé existuje další, užší klasifikace), udání kontextové zapojenosti a hloubkového slovosledu a nesou gramatické informace, jejichž skladba závisí na hloubkovém slovním druhu (vid, stupeň adjektiva, modalita, rod, negace, iterativnost děje, zdvořilost, ...). Uzel hlavního slovesa má určenou i větnou modalitu (oznamovací, tázací...). Navíc jsou určeny koreference – odkazy (např. zájmen) k objektům dříve zmíněným v textu.

Každý uzel má určené tektogramatické lemma, které ale někdy neodpovídá teorii z implementačních důvodů.