

Anotování aktuálního členění věty v Pražském závislostním korpusu

Kateřina Veselá, Jiří Havelka

ÚFAL/CKL Technical Report TR-2003-20

ISSN 1214-5521

Prosinec 2003

Rádi bychom poděkovali prof. Evě Hajičové a prof. Petru Sgallovi za teoretické vedení při přípravě manuálu pro anotování aktuálního členění a vyjádřili svůj vděk za jejich laskavost a trpělivost.

Výzkum, v jehož rámci vznikla tato technická zpráva, se uskutečnil za finanční podpory grantu MŠMT ČR LN00A063 a výzkumného záměru MSM113200006.

Obsah

1. Úvod	5
2. Úvodní komentář k manuálu pro anotaci aktuálního členění	7
2.1. Základní pojmy	7
2.2. Hlavní zásady	7
2.2.1. Kontextová zapojenost	8
2.2.2. Výpovědní dynamičnost.....	8
2.3. Pomocné instrukce	9
2.3.1. Povrchový slovosled	9
2.3.2. Typ závislosti	11
2.3.3. Nadvětné vztahy	13
2.3.4. Větná modalita	13
MANUÁL PRO ANOTACI AKTUÁLNÍHO ČLENĚNÍ.....	15
3. Vývoj manuálu pro anotaci aktuálního členění	31
3.1. Kvaziohnisko.....	31
3.2. Pořadí v základu	32
3.3. Pořadí v nominální skupině.....	33
3.4. Neprojektivní konstrukce	34
3.5. Rematizátory	34
3.6. Aktuální členění otázky.....	35
3.7. Drobnější změny	37
4. Výhledy	39
4.1. Kontrastivní téma	39
4.2. Zpracování složených predikátů a klasifikace neprojektivit.....	39
4.3. Rematizátory v nominální skupině.....	40
4.4. Vertikální členění textu	40
4.5. Podřadné souvětí	40
5. Technické řešení v Pražském závislostním korpusu	43
5.1. Závislostní stromy	43
5.1.1. Podmínka projektivity závislostních stromů.....	43
5.1.2. Algoritmus pro zprojektivnění/kontrolu projektivity.....	44
5.2. Programové nástroje pro modifikaci stromové struktury.....	45
5.2.1. Datová reprezentace závislostních stromů v editoru stromů TrEd	46
5.2.2. Makra pro editor stromů TrEd	46
5.2.2.1. Makra pro přiřazování hodnoty atributu tfa	46
5.2.2.2. Makra pro modifikaci pořadí uzlů/podstromů	47
Literatura	49

1. Úvod

Cílem této zprávy je především uvést poslední verzi manuálu pro značkování aktuálního členění v Pražském závislostním korpusu. K samostatnému pojednání této části Manuálu pro tektogramatické značkování nás vedl zejména fakt, že právě značkování aktuálního členění na hloubkové větné struktuře, typické pro Funkční generativní popis, není příliš rozšířeným jevem, přestože si jako významný syntaktický činitel zaslouží zvýšenou pozornost. Kromě toho se během posledních dvou let práce ukázalo, jak rozsáhlý je komplex jevů souvisejících s aktuálním členěním větným a do jaké míry má vliv na ostatní vrstvy značkování. Z pohledu aktuálního členění se mnohé diskutované problémy jeví v jiném světle, a proto se chceme pokusit podat tento pohled co nejkomplexněji. Manuál je přitom nejen výstupem, ale i východiskem naší práce – pokyny, které vznikají z potřeby upravit nově se vyskytnuvší jevy, zároveň umožňují prověřit hypotézy týkající se těchto jevů a podrobit je dalšímu zkoumání. Protože formulace v samotném manuálu jsou často kondenzovanými výstupy delších diskusí, připojujeme před manuál (2. kapitola) úvodní komentář. Ve třetí kapitole popisujeme, jak vznikaly některé pasáže manuálu a co nás k současným formulacím vedlo, čtvrtá kapitola pak popisuje nejdůležitější problémy, které je v anotaci aktuálního členění věty třeba v budoucnu ještě dopracovat. Poslední kapitola čtenáře stručně seznámí s technickým řešením anotování aktuálního členění v Pražském závislostním korpusu a s teoretickými otázkami s tím spojenými.

Protože samotný manuál pro anotaci aktuálního členění byl zpracován jako samostatný dokument s vlastním podrobným členěním a křížovými odkazy, nezačleňujeme jej kvůli přehlednosti do členění technické zprávy. Odkazy vyskytující se v samotném manuálu odkazují pouze v rámci manuálu, odkazy v textu ostatních částí technické zprávy na jednotlivé oddíly manuálu jsou uvedeny v podobě písmena M následovaného číslem příslušného oddílu (např. M 2.2.3.).

2. Úvodní komentář k manuálu pro anotaci aktuálního členění

Pokyny pro anotátory aktuálního členění obsažené v manuálu nejsou rovnocenné – k základním pokynům, které jsou závazné (oddíl Hlavní zásady), je připojen oddíl pomocných pokynů, které mohou být anotátorům vodítkem při obtížnějších situacích pro rozhodování (Pomocné instrukce). Tyto instrukce byly přidány zejména proto, že samotná kontextová zapojenost není přesně definovatelná a práce s kontextem je ovlivněna mnoha faktory, které se formování textu účastní. Na začátek manuálu byl také připojen oddíl Základní pojmy, který uvádí uživatele manuálu do termínů, které jsou dále v textu použity. Jednotlivé oddíly okomentujeme postupně, podle pořadí, ve kterém se v manuálu vyskytují.

2.1. Základní pojmy

V části Základní pojmy jsou vysvětleny především pojmy základ, ohnisko a vlastní ohnisko, které v naší teorii používáme. Základ je kontextově zapojená část věty (bráno v její lineární podobě). Jedná se o výrazy, vyskytující se zpravidla na začátku věty (před slovesem), které uvádějí větu do předchozího kontextu – svým významem navazují na výrazy, které se už v textu vyskytly, na skutečnosti z nich odvoditelné nebo na skutečnosti, u kterých se předpokládá, že jsou recipientovi známé (M 1.2.3).

Specifický způsob zapojenosti je navazování pomocí kontrastivního tématu, kdy je základ nové věty stavěn do kontrastu s nějakou skutečností známou z předchozího textu (M 1.2.4.).

Ohnisko je pak část věty, která uvádí „nové informace“ - je sdělně závažnější než základ (M 1.2.1.). Vlastní ohnisko je pak ta část ohniska, která je sdělně nejzávažnější a v mluvené podobě věty nese intonační centrum (M 1.2.2.). Tyto pojmy jsou v oddíle manuálu definovány v tektogramatickém stromě.

V oddílu je také uvedeno, jakých hodnot kontextové zapojenosti mohou jednotlivé uzly nabývat (M 1.1.).

2.2. Hlavní zásady

Oddíl Hlavní zásady obsahuje závazné pokyny pro anotátory, jak zachycovat aktuální členění na tektogramatické rovině. Naše anotace zachycuje dva důležité jevy aktuálního členění – kontextovou zapojenost a výpovědní dynamičnost.

2.2.1. Kontextová zapojenost

Kontextová zapojenost je ve značkování zachycena pomocí atributu TFA (Topic-Focus Articulation), který nabývá hodnot T (kontextově zapojený), F (kontextově nezapojený) a C (kontrastivně zapojený). Vodítkem pro přiřazení jednotlivých hodnot je zde význam daného uzlu a jeho vztah ke kontextu. Zapojenost určitého uzlu do kontextu není ovšem jednoduše definovatelná ze dvou hlavních důvodů. Za prvé Funkční generativní popis pojímá kontext široce – zahrnuje nejen bezprostřední verbální kontext („kontext“), ale i širší kontextové vrstvy, zahrnující veškeré sdílené či „obecně známé“ informace, jejichž sdílení může být podmíněno situací, kulturou, jinými texty a dalšími faktory. Za druhé i zapojenost výrazu do takto pojatého kontextu může mít nejrůznější podoby – od opakování a koreference přes vyplývání z většího textového úseku a odvozenost ze situace až po tak složité významové postupy jako např. metafora. Není proto jiné cesty než ponechat konkrétní rozhodnutí o kontextové zapojenosti na jazykovém povědomí anotátora. Vzhledem k tomu, že člověk dokáže spontánně tvořit věty se správným slovosledem, věříme, že dokáže i poměrně dobře intuitivně rozpoznat kontextovou zapojenost (viz M 2.1.).

2.2.2. Výpovědní dynamičnost

Výpovědní dynamičnost je ve stromech reprezentována tzv. hloubkovým slovosledem (M 2.2.). Nejobecnější pravidlo hloubkového slovosledu je řazení zapojených uzlů vlevo a nezapojených vpravo od řídicího uzlu. V jednotlivých patrech stromu by mělo pořadí uzlů odrážet vzrůstající výpovědní dynamičnost, přičemž nezapojené uzly jsou vždy dynamičtější než zapojené. Je ovšem také třeba upravit řazení sesterských zapojených a nezapojených uzlů mezi sebou. Toto řazení je přirozeně odlišné ve skupině verbální a nominální.

V nezapojené části verbální skupiny (ohnisku) by měla výpovědní dynamičnost odpovídat tzv. systémovému pořadí (Sgall, Hajičová, Panevová 1986). Abychom tuto hypotézu mohli ověřit, necháváme nezapojené uzly pod slovesem v souladu s povrchovým slovosledem – jen tak lze zjistit případné odchylky a zkoumat jejich příčiny. Kontextově zapojené uzly systémové pořadí odrážet nemusí, protože jejich výpovědní dynamičnost je dána užitím v daném kontextu. Jednotlivé skupiny uzlů, které mají v kontextově zapojené části věty z hlediska aktuálního členění odlišné funkce, řadíme podle předem určeného pořadí. Předpokládáme, že výpovědní dynamičnost je signalizována funkcí v rámci aktuálního členění, např. funkcí kulisy, centra tématu nebo nevyjádřeného uzlu. Jednotlivé funkce jsou pak řazeny podle vzrůstající výpovědní dynamičnosti, což v základu znamená přibližně podle klesající sdělné závažnosti (M 2.2.2.). Nejlevěji z tematických uzlů je tedy centrum základu a následují méně sdělné závažné uzly až po kulisy, popř. kontextově zapojené sloveso. (Ještě nalevo od něj jsou pak uzly, které se neúčastní větné struktury – vokativy, navazovací a modální částice).

V nominální skupině je pořadí jednotlivých skupin uzlů také určeno. Je zde nutné upravit hloubkový slovosled především proto, že povrchový slovosled je poměrně silně gramatikalizovaný. Těsnost atributů je signalizována velmi často gramatickou formou, a proto podle ní upravuje řazení atributů i manuál. V rámci jednotlivých skupin se opět držíme povrchového slovosledu, který je v těchto skupinách volný a tedy významově relevantní.

Výjimky ze slovosledného uspořádání jsou zpracovány v M 2.2.1., komentář viz v 3.1. a 3.5.

2.3. Pomocné instrukce

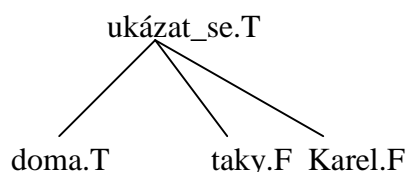
Jak bylo zmíněno výše, není pojem kontextové zapojenosti přesně definován, a proto může být rozhodování při evaluaci pro anotátory poměrně obtížné, zejména některých okrajových případech. Proto je v manuálu zařazen obsáhlý oddíl Pomocných instrukcí, který obsahuje další vodítka, která mohou anotátorům při evaluaci kontextové zapojenosti pomoci. Jde hlavně o silné tendence, které čeština vykazuje, a vliv čtyř důležitých faktorů – povrchového slovosledu (M 3.1.), významu uzlu (M 3.2.), nadvětných vztahů (M 3.3.) a větné modality (M 3.4.).

2.3.1. Povrchový slovosled

Slovosled je v češtině nejdůležitějším prostředkem pro vyjadřování aktuálního členění. Naznačovat hranici mezi základem a ohniskem výpovědi je funkcí slovesa, určité výrazy stojí na povrchu před slovesem v bezpříznakových případech právě proto, že jsou kontextově zapojené. Pokud stojí před slovesem kontextově nezapojený výraz, jde většinou o tzv. subjektivní pořadí, při němž je ohnisko z pragmatických důvodů kladeno na začátek věty (viz př. (2)). Jde ovšem o silně příznakový jev, objevující se většinou v mluvených textech, kde je dostatečně zřetelně signalizován polohou intonačního centra (M 3.1.2.). V psaném textu jej poznáme podle toho, že na konec věty jsou řazeny kontextově zapojené výrazy a samotné sloveso je většinou textově zapojené. Tyto případy anotujeme tak, že v hloubkové struktuře posuneme veškeré uzly do jejich nepříznakových pozic podle pokynů v Hlavních zásadách. Ve větách, kde vlastní ohnisko stojí v bezpříznakové pozici (na konci věty), jsou tedy výrazy stojící před slovesem téměř vždy kontextově nebo kontrastivně zapojené (M 3.1.1.). Podobně silná je tendence klást vlastní ohnisko na úplný konec věty, kde je také bezpříznakově umístěno intonační centrum (M 3.1.1). Výrazy kontextově nezapojené mají analogicky tendenci stát za slovesem, před vlastním ohniskem. Výjimkou jsou případy, kdy je sloveso kontextově zapojené, nebo kdy stojí na povrchu na „druhé pozici“ ve větě - ne na hranici základu a ohniska, ale za prvním větným členem, a další složky základu stojí za ním (viz př. (1)).

Příklady:

- (1) Prezident.F Klaus.T/C odjel.F včera.T na zahraniční.F nástěvu.F.
 (2) Taky.F KAREL.F se doma.T ukázal.T.



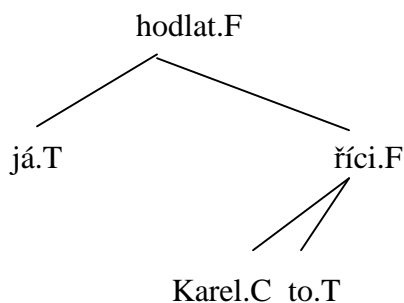
Povrchový slovosled je ovšem na rozdíl od tektogramatické struktury lineární, což s sebou nese určité problémy. Některé výše uvedené tendence například platí pouze pro uzly, závislé přímo na slovese, před slovesem se mohou i v bezpříznakové podobě věty vyskytnout kontextově nezapojená rozvití členů základu. Není tedy pravda, že všechny uzly, které stojí na povrchu před slovesem, musí být kontextově zapojené, protože rozvití kontextově zapojených uzlů jsou před sloveso předsunuty zároveň s nimi bez ohledu na to, zda jsou samy kontextově zapojené. V nominálních skupinách samotných je pak slovosled gramatikalizovaný a nemůžeme se podle něj vždy orientovat.

V některých případech dochází k jevu, kdy je určitý výraz v povrchovém slovosledu posunut na místo, kde by při přepisu na tektogramatickou rovinu způsobil porušení projektivity tektogramatického stromu (viz 5.1.1.). Zpracování takového typu vět je v manuálu věnován celý oddíl. Zatím uvažujeme o několika jevech, kterými může být neprojektivita motivována. Může být způsobena použitím kontrastivního tématu, které bylo na povrchu posunuto zcela doleva (viz př. (3)), nebo se jedná o kvaziohnisko, které je vlevo od řídicího slova, přestože je na něm závislé vlastní ohnisko, umístěné zcela doprava (viz př. (4)). Tyto posuny jsou příznakové a pragmaticky motivované, v hloubkové struktuře by tedy měly být v projektivní poloze (M 3.1.7.). U kontrastivních uzlů je nám pak povrchová neprojektivita pomocným signálem pro přiřazení kontrastivní zapojenosti.

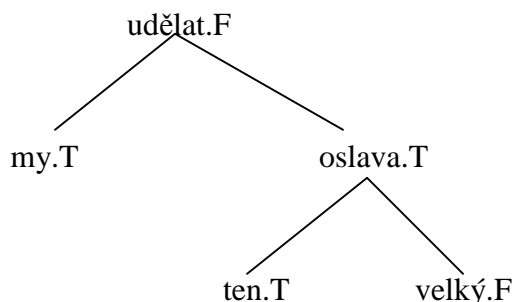
Výraz může být také umístěn vlevo ve větě z fonetických důvodů (chová se jako příklonka – viz př. (5)).

Příklady:

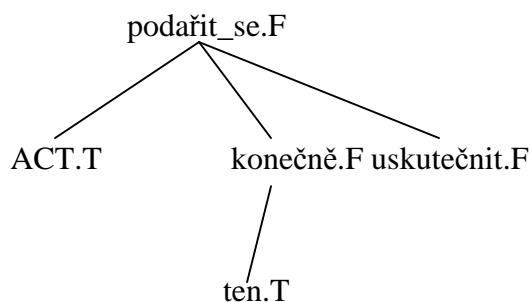
- (3) (Alena to ještě neví, ale) Karlovi.C to.T hodlám.F říct.F.



- (4) Musíme tu.T oslavu.T udělat.F velkou.F.



(5) Konečně.F se to.T podařilo.F uskutečnit.F.



Při práci jsme ovšem narazili i na další případy, kdy významová motivace přesunu není zřejmá a v některých z nich se dokonce jedná o bezpříznakové znění věty. U většiny těchto případů je motivace neprojektivity formální – jde o základ věty, jejíž predikát je složený z několika jednotek a na jeho závislé (kontextově nezapojené) části závisí kontextově zapojený uzel. Pokud v těchto případech není předsunutý uzel kontrastivní, necháváme prozatím tyto uzly v neprojektivních polohách (komentář tohoto jevu viz v 3.4.).

2.3.2. Význam uzlu

Při anotaci kontextové zapojenosti nám může pomoci i význam uzlu, nejen syntaktický, ale i lexikální.

Existují typy uzlů, které jsou zpravidla kontextově zapojené. Sem patří především uzly, které odkazují ke skutečnostem jako ke známým, a jsou tak zapojeny do kontextu situačního nebo společných znalostí - veškeré deiktické výrazy (některá zájmena, příslovce jako *tam*, *tady*, *tudy*, *tehdy*), nebo výrazy, které jsou odvozeny od známých skutečností (*včera*, *zítra*, *pozítrí*) (M 3.2.1.). Výjimkou jsou zde samozřejmě případy, kdy tyto uzly stojí ve vlastním ohnisku nebo v kontrastu. V této skupině, která je poměrně široká, samozřejmě existují typy, které tíhnou k zapojenosti silněji (ukazovací zájmena) a slaběji (osobní zájmena). Skupinou uzlů, které jsou vždy kontextově zapojené, jsou ty částice, které signalizují v textu návaznost na předchozí kontext – uzly typu PREC (*tedy*, *proto*, *však*, *tudíž*, *totiž* a další). Zapojovat větu do textového kontextu je jejich primární funkce (M 3.2.1.).

U některých pádů osobních zájmen existují v češtině dva tvary – silné (*tobě, tebe*) a slabé (*tě, ti*), z nichž slabé se vyskytují pouze v případech, kdy je zájmeno kontextově zapojené (M 3.2.5.).

Ke kontextové zapojenosti tíhnou také substantiva, která řídí číslovky - většinou jde o názvy jednotek, které jsou obecně známé a jsou ve větě vyjádřeny z gramatických důvodů (u některých vět mohou být na povrchu i vynechány), nebo se tu opakují výrazy známé z kontextu, které jsou číslovkou dále specifikovány (M 3.2.4.).

Analogicky některé skupiny uzlů tíhnou k tomu, být používány jako kontextově nezapojené. Jsou to silné tvary zájmen (M 3.2.6.). Dále výrazy, které označují něco neurčitého, neznámého (neurčitá zájmena, číslovky a příslovce – *někdo, několik, někde* apod.). Ke kontextové nezapojenosti tíhne také většina atributů, protože jejich funkcí je specifikovat řídící substantiva. Zvláštním případem jsou tzv. slovesné přívlastky – způsobová příslovce, která specifikují děj, vyjádřený slovesem (uzly s funktorem MANN).

Dále jsou v oddíle zpracovány některé zvláštní typy uzlů. Prvním z nich jsou uzly, které se v povrchové realizaci věty nevyskytují, jsou pouze na tektogramatické rovině. Na povrchu nejsou tyto uzly vyjádřené právě proto, že jsou považovány za vyvoditelné z kontextu. Jedná se o výrazy, které byly již v textu zmíněny, nebo jsou signalizovány v gramatických kategoriích jiných slov (podmět), nebo o ustálené konstrukce, kde se některý člen valenčního rámce na povrchu nevyjadřuje (např. všeobecný konatel). Právě proto, že se předpokládá jejich odvoditelnost z kontextu, jsou anotovány jako kontextově zapojené a jejich pořadí je stanoveno v oddíle o řazení uzlů v zapojené části věty.

Uzly s funktorem DENOM se vyskytují v nevětných konstrukcích, většinou jde o titulky, seznamy apod. Protože tyto konstrukce nebývají navázány na kontext, většinou dostávají uzly DENOM hodnotu F (M 3.2.11.).

Uzly typu PAR jsou řídící uzly vsunutých vět, nebo pouze členů (M 3.2.12.). Hodnota kontextové zapojenosti se určuje podle významu (zpravidla je to F), jejich pořadí ovšem zůstává shodné s povrchovým, protože se jedná o zcela nezávislé struktury, které nemají své místo ve struktuře věty (jedná se o výjimku ze závazného pravidla – viz M 2.2.1.).

Dalšími nezávislými členy jsou oslovení (uzly typu VOCAT) – ty jsou hodnoceny jako kontextově zapojené, protože odkazují k adresátovi textu, který je součástí situačního kontextu. Oslovení se většinou vyskytují na začátku věty, kam jsou také v hloubkové struktuře řazena. na první místo v základu jsou pak posouvána, i když na povrchu leží jinde (M 3.2.13.).

Velmi specifické je chování tzv. rematizátorů (M 3.2.10.). To jsou částice, jejichž funkcí je signalizovat hranici ohniska, popř. vlastního ohniska. Jejich umístění v povrchovém pořadí je poměrně volné, podle toho, kterou část věty „rematizují“ - signalizují, že není kontextově zapojená. O této části věty platí, že je v dosahu rematizátoru. V tektogramatickém stromě jsou to všechny jeho sesterské uzly stojící

napravo od něj. Řídící uzel je v dosahu rematizátoru tehdy, když se jedná o sloveso, v ostatních případech je rematizátor levou sestrou prvního uzlu, který má v dosahu. Pokud je tedy v dosahu rematizátoru řídící uzel, má výjimku ze závazné zásady, a přestože má F (rematizátor uvozující ohnisko je sám členem ohniska), řadí se doleva od řídícího uzlu. V ostatních případech se řadí vpravo od řídícího uzlu a vlevo od prvního sesterského uzlu, který má v dosahu. v některých případech se může rematizátor vyskytnout i v základu – tehdy předpokládáme, že signalizuje přítomnost kontrastivního tématu. V takových případech má rematizátor T a píše se vlevo od kontrastivního tématu. Problematické je umístění rematizátoru v nominální skupině (viz 3.5.).

2.3.3. Nadvětné vztahy

V předposledním oddíle manuálu se hovoří o vztazích v textu, přesahujících rámec věty. Probíráme zde zpracování souvětí souřadného, ale i souvětí podřadného (přestože u něj ve skutečnosti o nadvětné vztahy nejde; domníváme se však, že problematika souvětí by měla být řešena na jednom místě – M 3.3.1.), a také jevů jako jsou titulky (M 3.3.2.) nebo přímá řeč (M 3.3.1.2.). Obecně platí, že souvětí souřadné je pojímáno a anotováno jako více vět (těch, které jsou koordinovány) a souvětí podřadné je anotováno jako jedna věta. Přejít mezi oběma typy není ovšem z hlediska aktuálního členění příliš ostrý, proto je zavedena výjimka pro některá souvětí, která, ač jsou formálně podřadná, se z hlediska významu chovají jako souřadná (jedná se zde o chování některých závislých klauzí příslovečného typu – diskusi viz v 4.5.).

2.3.4. Větná modalita

Poslední oddíl manuálu (M 3.4.) se zabývá aktuálním členěním otázky. Anotování otázky se řídí v podstatě stejnými pravidly jako anotace oznamovací věty, pouze orientace v povrchové realizaci věty je poněkud odlišná. Jde o případy inverzního slovosledu u otázek zjišťovacích a tázací výrazy u otázek doplňovacích. Sloveso, které stojí u zjišťovacích otázek s inverzním slovosledem na začátku věty a tázací výraz, který stojí na téže pozici u otázek doplňovacích, jsou anotovány jako nejméně dynamické členy ohniska, pokud nejsou ve vlastním ohnisku a pokud je sloveso kontextově nezapojené (podrobnosti viz v 3.6.).

MANUÁL PRO ANOTACI AKTUÁLNÍHO ČLENĚNÍ

1. Základní pojmy

1.1. Jednotlivé uzly mají v atributu TFA hodnoty:

- T - kontextově zapojený výraz, primárně část základu („topic“),
- C - kontrastivní část základu, kontrastivní téma („contrastive topic“),
- F - kontextově nezapojený výraz, primárně část ohniska („focus“).

Tyto hodnoty budou předvyplněny automatickou procedurou, ovšem pouze jako východisko pro ruční analýzu - definitivní přiřazení příslušných hodnot uzlům provádí anotátor.

1.2.1. Pod pojmem ohnisko rozumíme

- (a) uzly s F přímo závislé na slovese a sloveso, pokud má F;
- (b) hlouběji závislé uzly s F, které mají nadřazený uzel s T, pokud jeden z nich je nositelem IC věty (IC=intonační centrum - slovo/takt, který nese v mluvené podobě výpovědi klesavý „větný“ přízvuk - v textu dále je vyznačeno verzálkami);
- (c) další uzly závislé na členech ohniska.

1.2.2. Vlastním ohniskem rozumíme nejdynamičtější uzel (popř. substantivní skupinu, viz Pozn. 1.), výraz nesoucí v mluvené podobě výpovědi intonační centrum, který je v hloubkovém slovosledu vždy zcela vpravo, i když stojí v povrchovém pořadí na jiné pozici (tzv. subjektivní pořadí).

Pozn. 1. Pokud je vlastním ohniskem kontextově nezapojená část substantivní skupiny, může mít přízvuk na posledním slovu, i když to není jejím nejdynamičtějším členem.

1.2.3. Pod pojmem základ rozumíme

- (a) uzly s T přímo závislé na slovese, kterým není podřízeno vlastní ohnisko, a sloveso, pokud má T;
- (b) další uzly závislé na členech základu odlišných od kořene stromu.

1.2.4. Pod pojmem kontrastivní téma rozumíme kontrastivní část základu, která zahrnuje

- (a) uzel s hodnotou C (kontrastivně zapojený), který nese fakultativní stoupavý „kontrastivní“ přízvuk (viz Pozn. 1.)
- (b) uzly, které jsou mu podřízeny.

2. Hlavní zásady

2.1. Přiřazení hodnot T, F, C

Hodnoty T, C a F přiřazujeme podle platnosti slov v daném kontextu. Kontextem přitom rozumíme nejen okolí věty v daném textu („kotext“), ale i širší souvislosti tohoto textu (intertextové, situační, kulturní apod.).

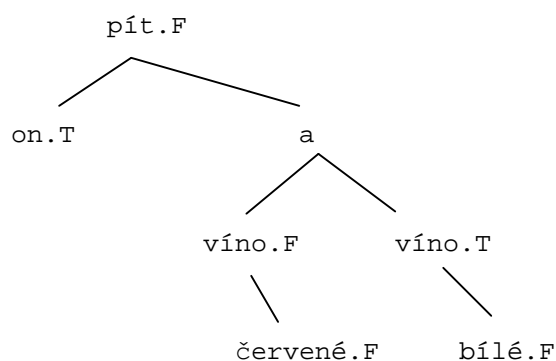
2.1.1. Hodnota T je přiřazena uzlům, které reprezentují opakující se (i když ne doslova) jednotky z předcházejícího textu (uvnitř věty i před ní) nebo nějak související s širším kontextem (věci „evidentní“ ze situace nebo obecného povědomí); jde především o případy koreference ap. (i když jde o sloveso).

2.1.1.1. Do této skupiny patří také uzly, které se v tektogramatickém stromu doplňují (Emp.PRED, uzly typu Gen a Cor, Idph, elidované uzly...).

2.1.1.2. Výjimky jsou u negace, koordinace a binárních relací typu „z...na...“, kde je vypuštěn první výskyt substantiva - ten je pak F nebo C, zatímco vyjádřený druhý výskyt je T.

Příklad 1:

Pít.F červené.F (víno.F) a bílé.F víno.T.



2.1.2. Hodnota C se přiřazuje kontrastivní části základu, která má většinou tyto vlastnosti:

2.1.2.1. Předpokládá výběr z množiny alternativ (množina nemusí být v textu vyjmenována - pak je kontrastivní téma neodvoditelné z bezprostředního verbálního kontextu).

2.1.2.2. V mluvené podobě výpovědi nese fakultativní stoupavý („kontrastivní“) přízvuk.

2.1.2.3. U zájmen je vyjádřena silným tvarem.

2.1.2.4. Na povrchu stojí v iniciální pozici ve výpovědi.

Příklady 2-4:

Jedině.T s úspěšnými.F vzory.C se můžeme poměřovat.F.

Jemu.C to.T Martin.T nedal.F.

Janu.C Marie.T neviděla.F.

2.1.3. Hodnota F je přiřazena uzlům, které reprezentují v textu dosud neznámá („nová“) fakta nebo souvislosti, tj. informace, které nejsou

odvoditelné z kontextu, a která nejsou (nemohou být) nositeli kontrastivního přízvuku.

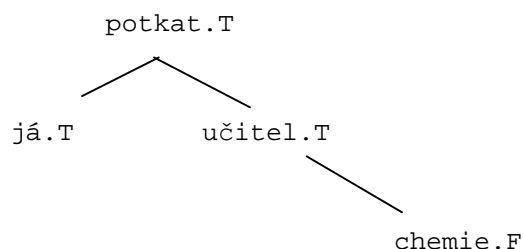
2.2. Pořadí uzlů v tektogramatické struktuře

2.2.1. T i C vždycky závisí vlevo od řídicího uzlu a F závisí vpravo od řídicího uzlu, kromě těchto výjimek:

2.2.1.1. Kvaziohnisko je kontextově zapojené slovo, jemuž je podřízeno¹ vlastní ohnisko - má T, ale je vpravo.

Příklad 5:

(Kterého učitele jsi potkal?)² Potkal.T jsem učitele.T chemie.F



2.2.1.2. Rematizátor může mít F, ale být vlevo od řídicího uzlu (viz 3.2.9.1.).

2.2.1.3. Uzel typu PAR má zpravidla F, ale zůstává na stejné pozici, jako v povrchovém slovosledu i pokud je vlevo od řídicího uzlu

2.2.1.4. Predikáty některých typů závislých klauzí mohou mít F a přitom stát vlevo od hlavního predikátu (viz 3.3.1.3.)

2.2.1.5. V koordinačních skupinách, kde je první výskyt substantiva elidován, má F, ale leží vlevo od koordinačního uzlu (viz Příklad 1)

2.2.2. Pořadí jednotek v kontextově zapojené části není libovolné. Předpokládá se toto uspořádání:

(a) Nejvíce vlevo se umístí uzel vokativu (VOCAT), pokud je ve výpovědi obsažen,

(b) za něj se umístí uzly s funktorem PREC, dále uzly typu ATT,

(c) rematizátor, pokud má v dosahu kontrastivní uzel,

(e) za něj uzly kontrastivního tématu,

(f) dál následují vyjádřené tematické uzly s výjimkou typu (h) níže,

(g) dále elidované uzly,

(h) uzly odkazovacího zájmena „on“,

(i) kulisy, tj. místní a časová doplnění v T, která nejsou C,

(j) sloveso, pokud je kontextově zapojené.

1 uzel A je podřízen uzlu B, pokud je každý další uzel mezi ním a uzlem B ve vztahu přímé nebo nepřímé závislosti k uzlu B (vztah podřízenosti je tranzitivním uzávěrem vztahu (přímé) závislosti

2 v závorkách je uváděn předchozí kontext

Pokud je uzlů v některé skupině více, řadí se v pořadí shodném s povrchovým slovosledem.

Pozn. 2. Výrazy jako *u nás, minulou sobotu* jsou zpravidla kulisami, ale nemusí jimi být (viz Příklad 6).

Příklad 6:

(Včera jsem byl v knihovně.) Dnes.C (kontrastivní téma) tam.T (kulisa) nepůjdu.F.

Přísllovečný výraz může být i v ohnisku, pak ovšem není kulisou (viz Příklad 7).

Příklad 7:

Půjdu.T/F tam.T až.F dnes.F.

2.2.3. Pořadí v nominální skupině

2.2.3.1. Kontextově nezapojené atributy se řadí vpravo od řídicího uzlu zpravidla v tomto pořadí:

- (a) části ustálených spojení, uzly typu DPHR
- (b) uzly typu ID
- (c) valenční přívlasky (aktanty deverbativ), genitivní přívlasky (uzly typu APP a MAT)
- (d) shodné přívlasky (adjektiva, posesiva, číslovky, u deverbativ MANN, EXT,) v opačném pořadí než v povrchovém slovosledu
- (e) nevalenční (tj. volná fakultativní) neshodná doplnění (převážně místní a časová určení)
- (f) nerestriktivní vztažné klauze.

2.2.3.2. Pokud je v nějaké ze skupin (a)-(f) více členů, řadí se podle povrchového slovosledu; shodné přívlasky se přeřazují zleva do prava zpravidla v obráceném pořadí, než je povrchové.

2.2.3.3. Kontextově zapojené atributy stojí vlevo od svého řídicího uzlu; v jejich řazení se postupuje analogicky, ovšem zrcadlově, tj. v pořadí (f) - (a), se shodnými přívlasky podle povrchového slovosledu.

3. Pomocné instrukce

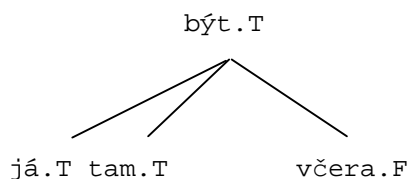
3.1. TFA ve vztahu k povrchovému slovosledu

3.1.1. Uzly, které závisí zleva na slovesu (určitém tvaru ve složeném tvaru slovesa, infinitivu, pokud není ve složeném tvaru slovesa), mají zpravidla T nebo C.

3.1.2. Vyjímkou jsou případy, kdy je vlevo od slovesa vlastní ohnisko a za ním následují zapojené uzly (tzv. subjektivní pořad - viz 1.2.2.). Vlastní ohnisko pak dostává F a řadí se úplně doprava.

Příklad 8:

(Ne, nebyl jsem tam v pátek.) VČERA.F jsem tam.T byl.T.



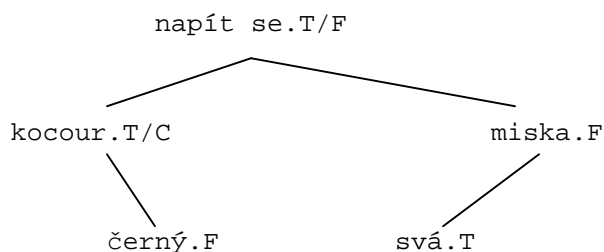
3.1.3. Uzel, který závisí na slovesu nejvíc doprava, má F (kromě výjimek uvedených zde v 3.1.2. - v těch zapojené uzly dostanou T a posouvají se doleva);

3.1.4. Uzly závislé na slovese, které leží v povrchovém slovosledu napravo od něj, mají F nebo T podle toho, zda se opakují nebo jsou vyvoditelné z předchozího kontextu; místní, časová a šíře okolnostní určení mají T, pokud je lze považovat za kulisu (viz. Pozn. 2.).

3.1.5. Hodnota hlouběji závislých uzlů nezávisí na jejich povrchovém postavení vpravo nebo vlevo od slovesa.

Příklad 9:

Černý.F kocour.T/C se napil.T/F ze své.T misky.F.



3.1.6. Shodné přívlastky se přeřadí v zrcadlovém pořadí ze své povrchové pozice nalevo od svého řídicího uzlu doprava od něj (viz 2.2.3.2.), pokud není kontextem dáno, že jsou T.

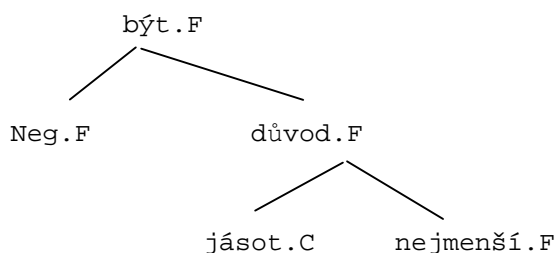
3.1.7. Zachycení struktur, které jsou na analytické rovině neprojektivní:

Zatím navrhujeme ve vzorovém i velkém souboru tato řešení

3.1.7.1. Uzel zavěšený neprojektivně zcela vlevo bude označen (kromě případů v bodech 3.1.7.2 a 3.1.7.3.) symbolem C(kontrast), bude se tedy předpokládat jeho kontrastivní užití (mělo by být schopné nést fakultativní stoupavý „kontrastivní„ přízvuk) a zavěsí se projektivně co nejlevěji.

Příklad 10:

K jáсотu.C není.F nejmenší.F důvod.F.



3.1.7.2. Uzel stojící „neprojektivně“ vlevo může dostat T, pokud je z kontextu zřejmé, že o kontrast evidentně nejde, což se předpokládá u těchto případů:

3.1.7.2.1. Uzel závisí na závislé části složeného predikátu (uzlu typu CPHR).

3.1.7.2.2. Uzel závisí na závislé části konstrukce s kontrolou řízené kvazimodálním nebo fázovým slovesem (popř. složeným predikátem s kvazimodálním nebo fázovým významem).

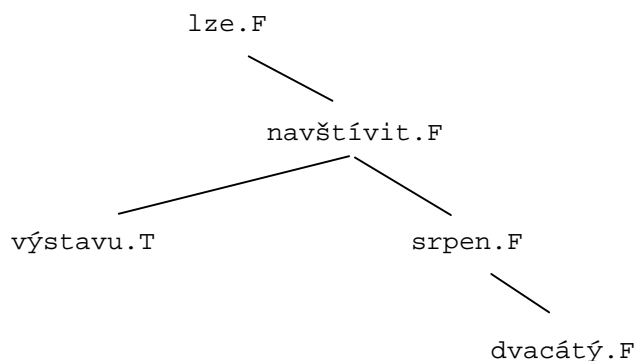
3.1.7.2.3. Uzel závisí na závislé části frazému se slovesem „být“ typu *je nutné, je třeba, je dobré* apod.

(viz připravované dodatky o složených predikátech, frazémeh a kontrole k Manuálu pro tektogramatické značkování)

V těchto případech zůstává uzel na původní pozici, neposouvá se doprava, protože jeho „neprojektivní“ postavení je zapříčiněno způsobem zpracování složených slovesných konstrukcí - uzel stojící vlevo se vztahuje k celému frazému a jeho povrchové postavení nemusí být motivováno kontrastivním užitím.

Příklad 11:

(V galerii V. Špály ode dneška vystavuje A. Born.) Výstavu.T lze.F navštívit.F do dvacátého.F srpna.F.

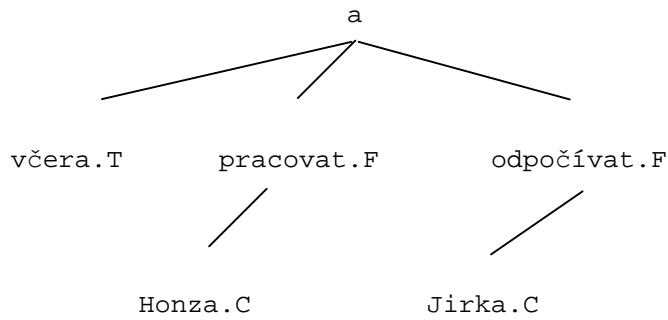


Pozn. 3. Jde o prozatímní řešení, které bude nahrazeno projektivními strukturami, jakmile budou zpracovány různé stupně frazeologických spojení. Soupis frazémů a kvazimodálních sloves, u kterých k této neprojektivitě dochází, se bude postupně pořizovat.

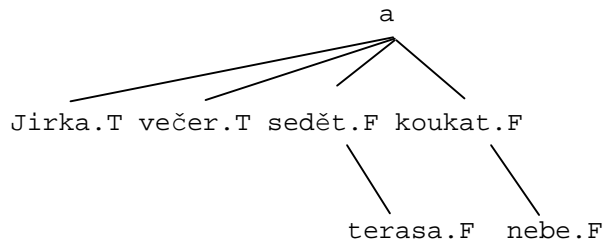
3.1.7.3. Uzel, který rozvíjí celou koordinační konstrukci, předchází před všemi uzly, které závisejí na jednotlivých částech této konstrukce (je-li zapojený), nebo za nimi následuje (je-li nezapojený);

Příklady 13-14:

Honza.C včera.T pracoval.F a Jirka.C odpočíval.F.



Večer.T seděl.F Jirka.T na terase.F a koukal.F na nebe.F.



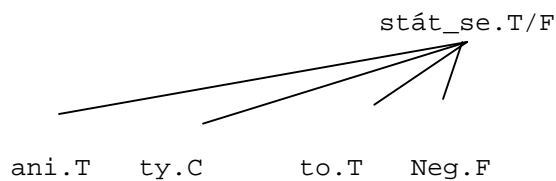
3.2. TFA a význam uzlu

3.2.1. Indexické výrazy - *já, my, teď, můj, náš, tady, letos, zítra* apod. - i výrazy s funktorem PREC (*proto, tedy* aj.) mají obvykle přiřazenu hodnotu T s těmito výjimkami:

3.2.1.1. Při kontrastivním užití mají C.

Příklad 15:

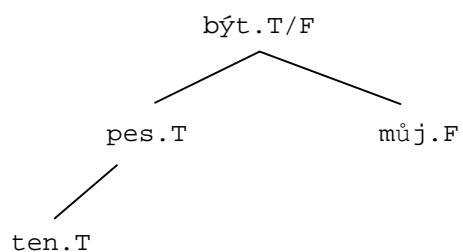
Ani.T tobě.C se to.T nestane.F.



3.2.1.2. Jako nositelé intonačního centra mají F.

Příklad 16:

Ten.T pes.T je.T/F můj.F.

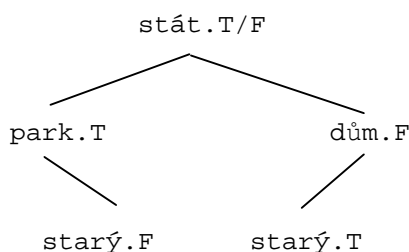


3.2.2. Obecná zájmenná slova - *někdo, něco, jednou, nějaký apod.* mají zpravidla hodnotu F.

3.2.3. Atribut vyjádřený adjektivem, substantivem, zájmenem (kromě indexických ap.) nebo číslovkou a závislý na substantivu má zpravidla přiřazeno F, i když je v povrchovém slovosledu vlevo od substantiva; hodnota T se přiřazuje, jen když se tento atribut opakuje nebo je z předcházejícího kontextu jasný:

Příklad 17:

Ve starém.F parku.T stál.T/F starý.T dům.F.



3.2.4. Hodnota T se přiřadí substantivům, která řídí základní číslovky ap. (řídící substantivum může být v těchto příkladech v pozici kvaziohniska).

Příklad 18:

Utratil.T za tu.T učebnici.T korun.T sto.F:

Výjimkami jsou:

3.2.4.1. Pokud jsou užitá kontrastivně, tj. např. koruny jsou v kontrastu k dolarům, léta k měsícům ap., substantivum dostává C.

Příklad 19:

U nás by to stálo šedesát.F korun.T; za dva.F dolary.C to ovšem v USA nedostaneš.

3.2.4.2. Pokud je substantivum ve vlastním ohnisku (např. když se zmiňuje v textu vůbec poprvé), dostává F.

3.2.5. Hodnota T je přiřazena slabým tvarům zájmen (*tě, ti, ho, mu, mi*) a jiným příklonkám, i tvarům *mě, mně, ji, jí, jim, jej, nás* atd.), pokud na nich není umístěn větný nebo kontrastivní přízvuk.

3.2.6. Silné tvary zájmen (*tebe, jemu* atd.) mají přiřazenu hodnotu F (pokud je na nich intonační centrum - viz 1.2.2.) nebo C (při kontrastivním užití).

3.2.6.1. Zájmeno za předložkou má vždycky silný tvar, takže může dostat i hodnotu T.

Příklady 20-23:

Pro něj.C to přinesu ZÍTRA.

Zítra to přinesu PRO něj.F.

Zítra to pro něj.T PŘINESU.

ZÍTRA to pro něj.T přinesu.

3.2.7. Symboly řídící koordinační spojení nemají T ani F a píšou se mezi částí koordinovaného spojení (viz Příklad 1).

3.2.8. Doplněný uzel je zavěšen vždy vlevo od svého řídícího uzlu a má index T (výjimky viz v 2.1.1.2. a 2.2.1.1.).

3.2.9. Neg a další rematizátory:

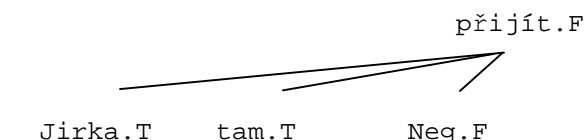
3.2.9.1. Rematizátor visí (kromě výjimek v 3.2.9.2.) vždy jako levá sestra prvního uzlu, který má v dosahu.

3.2.9.2. Pokud rematizátor závisí na predikátu, může být predikát v jeho dosahu. V těchto případech se řídíme následujícími pravidly:

3.2.9.2.1. Pokud je predikát v dosahu rematizátoru a je kontextově nezapojený, má rematizátor F, ale je vlevo (výjimka z pravidla o pořadí - viz 2.2.1.2.):

Příklad 24:

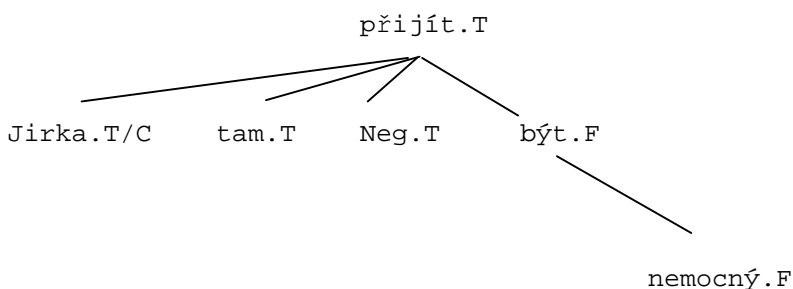
Jirka.T/C tam.T nepřišel.F.



3.2.9.2.2. Pokud je predikát v dosahu rematizátoru a je kontextově zapojený, má rematizátor T a je vlevo.

Příklad 25:

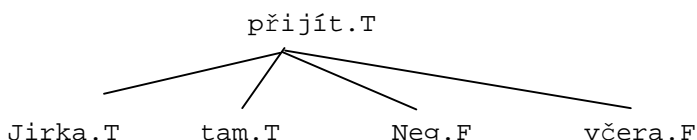
Jirka.T/C tam.T nepřišel.T proto, že je.F nemocný.F.



3.2.9.2.3. Jestliže predikát není v dosahu rematizátoru, má rematizátor F a je vpravo.

Příklad 26:

Jirka.T/C tam.T ne.Fpřišel.T včera.F, (ale až dnes.)

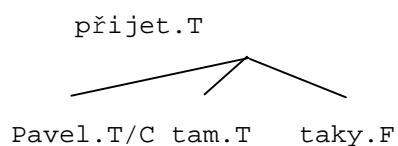


3.2.9.3. Pokud rozvíjí nominální frázi (právě.RHEM velký červený vůz), jsou v jeho dosahu všechny nezapojené výrazy, které za ním následují v povrchovém pořadí.

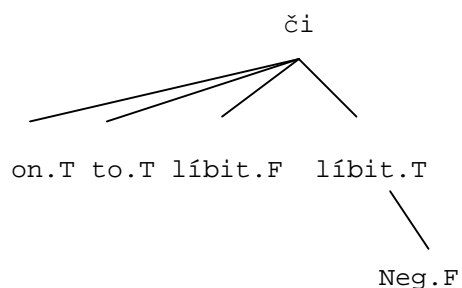
3.2.9.4. Je-li rematizátor nositelem intonačního centra, vytváří sám celé ohnisko.

Příklady 27-28:

Pavel.T/C TAKY.F přijel.T



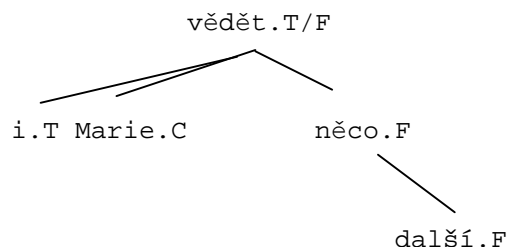
(Udělám to,)at se mu.T to.T líbí.F, či ne.F(líbí.T).



3.2.9.5. Pokud je v jeho dosahu slovo s hodnotou C, dostává rematizátor T a zůstává vlevo od uzlu, který má v dosahu.

Příklad 29:

I.T Marie.C ví.T/F něco.F dalšího.F.

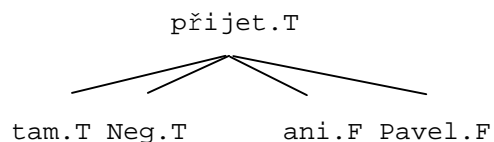


3.2.9.6. I ve složitějších případech se řídíme vždy dosahem rematizátoru a hodnotou kontextové zapojenosti uzlu/ů, který/é má v dosahu.

Příklady 30-31:

Ne.T_přijel.T ani.F PAVEL.F.

Ani.F PAVEL.F ne.T_přijel.T.

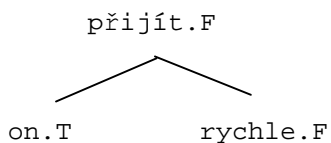


3.2.10. Volná doplnění

3.2.10.1. Příslovce s MOD a MANN má zpravidla F a přesouvá se doprava (i když stojí na povrchu vlevo, tzv. adverbální přívlástek).

Příklad 32:

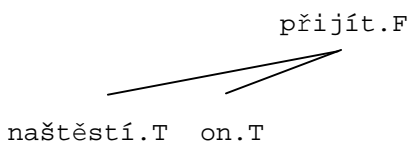
Rychle.F přišel.F.



3.2.10.2. ATT má zpravidla T.

Příklad 33:

Naštěstí.T přišel.F.



3.2.11. V nevětných konstrukcích má DENOM zpravidla F (ve větách s Emp.PRED může mít ovšem iniciální člen T nebo C!).

Příklady 34-35:

Škola.DENOM.F šachových.F koncovek.F

ale

Sparta.ACT.C (Emp.PRED.T)ve finále.F

3.2.12. Uzly typu PAR zůstávají na místě, které mají v povrchovém slovosledu, a zpravidla dostanou F.

3.2.13. Oslovení (VOCAT) dostávají zpravidla T (nejsou-li vlastním ohniskem nebo e kontrastu) a posouvají se na nejlevější pozici ve stromě.

Příklady 36-38:

Jirko.T, podej.F mi.T pero.F.

(Jirko, podej mi pero.) Honzo.C,(podej.T mi.T.) ještě.F zápisník.F.

3.3. Nadvětné vztahy

3.3.1. Pro souvětí platí především:

3.3.1.1. Souvětí souřadné

Každá koordinovaná klauze má své vlastní aktuální členění.

Příklady 39-41:

Tom.T přinesl.F knihy.F a pak.T Jirka.C odnesl.F noviny.F

Knihy.C odnesl.F a noviny.C přinesl.F.

Chodil.F sem.F a chodil.T tam.F, ale nenašel.F nic.F.

3.3.1.2. Přímá řeč je zpravidla dynamičtější než uvozovací věta (pokud v uvozovací větě není vlastní ohnisko - viz Příklad 44).

Příklady 42-44:

Jirka.T/C řekl.T/F: „Je.F dobře.F.“
„Je.F dobře.F.“, řekl.T/F Jirka.T/C.
Taky.F JIRKA.F řekl.T „Je.T dobře.T“.

3.3.1.3. V podřadném spojení klauzí se závislá obvykle nechá vlevo nebo vpravo podle povrchového pořadí a predikát závislé klauze dostane F v těchto případech:

3.3.1.3.1.1. Jde o obsahovou nebo příslovečnou „vedlejší větu“, která stojí za klauzí řídicí, popř. je vložena do její rematické části (kromě případů s kvaziohnskem).

3.3.1.3.1.2. Závislá klauze uvádí relativně samostatný nový děj a předchází řídicí klauzi z širší sémantických důvodů (časová nebo kauzální následnost dějů) – v takovém případě zůstává klauze vlevo od hlavního predikátu, i když má F.

3.3.1.3.2. Pokud je sloveso příslovečné klauze před slovesem klauze řídicí a je v kontrastivním postavení (má fakultativní stoupavý „frázový“ přízvuk), dostane toto sloveso C a zůstane vlevo.

Příklad 45:

Protože se program.T osvědčil.C, použili.F ho.T.

Při pochybnostech má toto řešení s C přednost před 3.1.3.1.

3.3.1.3.3. V ostatních případech dostává T.

Pozn. 4. Tatáž pravidla platí i pro nominalizace těchto klauzí

3.3.1.4. Sloveso v uvozovací klauzi (na kterém závisí DSP) má zpravidla T.

3.3.1.5. Má-li něco C ve vztažné klauzi, bude to vlevo od vztažného slova (to se posouvá doleva až na povrchu).

3.3.2. Pokud věta následuje za titulkem (pokud na něj přímo a zřetelně svým členěním nenavazuje), její kontextová zapojenost se určuje, jako by šlo o první větu textu.

Příklady 46-47:

(Dnešní zasedání sněmovny.) Na dnešním.T zasedání.T sněmovny.T předložili.F...

(Dnešní zasedání sněmovny.) Vládní poslanci předložili na dnešním.T zasedání.F sněmovny.F...

3.4. Otázka

3.4.1. Otázka zjišťovací

3.4.1.1. Výraz nesoucí v mluvené podobě IC otázky dostává F a posouvá se zcela vpravo, jakožto vlastní ohnisko otázky.

3.4.1.2. Sloveso se, pokud není vlastním ohniskem, posouvá mezi kontextově zapojenou a nezapojenou část výpovědi.

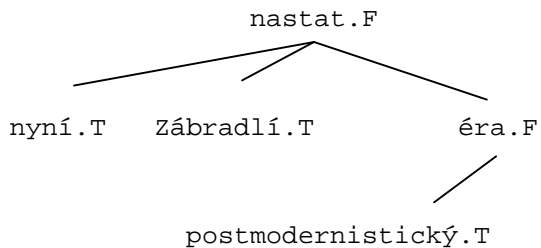
Kontextově zapojené výrazy dostávají T a posouvají se doleva od slovesa v pořadí shodném s pořadím v kontextově zapojené části oznamovací věty (viz 2.2.2.).

3.4.1.3. Výraz, který můžeme vyslovit s kontrastivním přízvukem a není zřejmě vyvoditelný z předchozího kontextu, dostává C a posouvá se úplně doleva. To platí i pro výrazy stojící vpravo od slovesa.

3.4.1.4. Nezapojené výrazy dostávají F a řadí se mezi sloveso a vlastní ohnisko v pořadí shodném s povrchové.

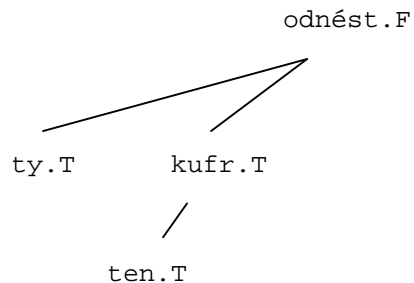
Příklady 49-50:

Nastane.F nyní.T na Zábradlí.T postmodernistická.T éra.F?



Pozn. Výraz „postmodernistický“ má T, protože je zmíněn v předchozím kontextu.

Odnesl.F jsi.F ten.T kufr.T?



3.4.2. Otázka doplňovací

3.4.2.1. Výraz nesoucí v mluvené podobě IC otázky dostává F a posouvá se zcela vpravo, jakožto vlastní ohnisko otázky. Pokud se IC nedá jednoznačně určit, vychází se z předpokladu, že vlastním ohniskem je tázací výraz.

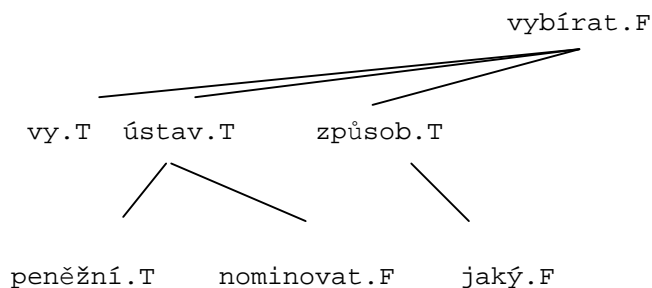
3.4.2.2. Tázací výraz se, pokud není vlastním ohniskem, posouvá v nezapojené části výpovědi těsně za sloveso.

3.4.2.3. Výraz, který stojí mezi tázacím výrazem a slovesem, který můžeme vyslovit s kontrastivním přízvukem a není přímo vyvoditelný z předchozího kontextu, má C a posouvá se zcela doleva. To platí i pro výrazy stojící vpravo od slovesa.

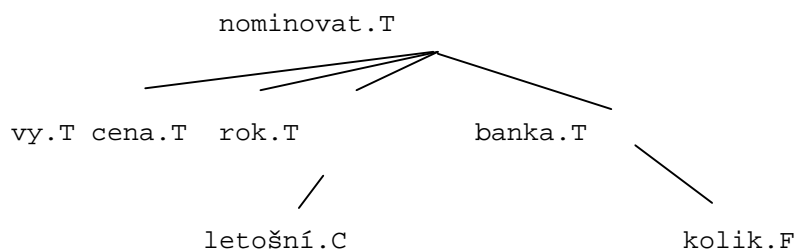
3.4.2.4. Ostatní výrazy ve výpovědi se řídí stejnými pravidly jako v oznamovací větě.

Příklady 51-53:

Jakým.F způsobem.T peněžní.T ústavy.T, které.T chcete.F na ocenění.T nominovat.T, vybíráte.F ?

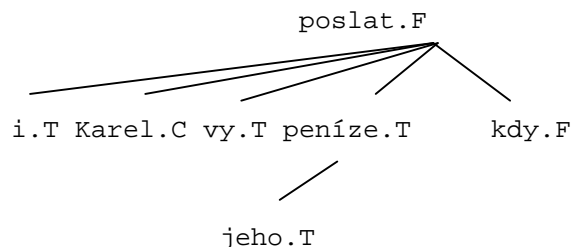


KOLIK.F bank.T jste.T v letošním.C roce.T na cenu.T nominovali.T?



Pozn. Výraz „banka“ je znám z předchozího kontextu, v uvedené větě jde o kvaziohnisko.

Kdy.F pošlete.F i.T Karlovi.C jeho.T peníze.T?

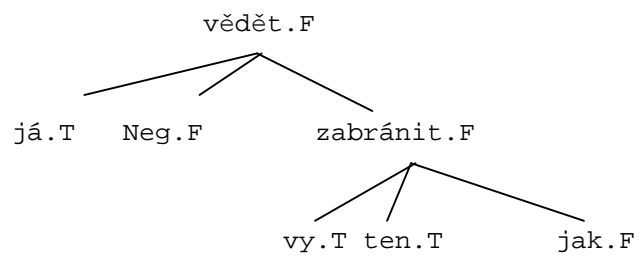


3.4.3. Nepřímá otázka

3.4.3.1. U většiny nepřímých otázek se postupuje stejně jako u otázky doplňovací.

Příklad 54:

Nevím.F, jak.F tomu.T zabráníte.F.



3.4.3.2. Nepřímé otázky uvozené výrazy typu „zda“, „jestli“ se chovají jako otázky zjišťovací.

3. Vývoj manuálu pro anotaci aktuálního členění

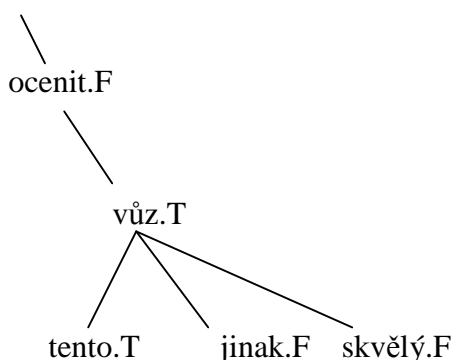
Výchozím materiálem pro manuál aktuálního členění byla třetí kapitola Manuálu pro tektogramatické značkování (Sgall, Hajičová 2001). Během dvouleté práce s tímto manuálem bylo třeba provést větší množství úprav, které si vyžádaly příklady jevů objevujících se zcela nově, nebo zpracovaných ve stávajícím manuálu nedostatečně. K těm nejpodstatnějším patřilo nové členění a uspořádání manuálu (viz kapitolu 2), rozšíření příkladového aparátu a dopracování teoretických problémů, jejichž výčet se pokoušíme níže uvést včetně důvodů, které nás ke změnám vedly.

3.1. Kvaziohnisko

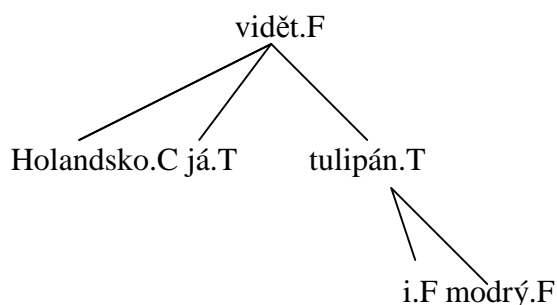
Kvaziohnisko je nový termín, který nahradil předešlé pojmenování proxy focus. To označovalo kontextově zapojený uzel vpravo od slovesa, kterému bylo podřízeno vlastní ohnisko. Předpokládalo se, že se vyskytuje především ve větách, ve kterých je hranice základu a ohniska posunuta dále za sloveso (viz příklad (8)). Příklady (6) a (7) ovšem ukazují, že uzel hodnocený jako kontextově zapojený, kterému je podřízeno vlastní ohnisko, nemusí poukazovat pouze na posun hranice mezi základem a ohniskem, a jako takový nemusí být podřízen jen kontextově zapojenému uzlu, ale i kontextově nezapojenému. Případy, kdy je vlastní ohnisko podřízeno kontextově zapojenému uzlu, se vztahují na veškeré konstrukce, v nichž je vlastním ohniskem uzel specifikující výraz známý z kontextu. V těch ovšem může být hranice mezi základem a ohniskem umístěna bezpříznakově před nebo těsně za slovesem. Pojem kvaziohnisko tedy rozšiřuje původní pojetí proxy focusu a značí libovolný kontextově zapojený uzel, pod nímž se ve stromě nachází vlastní ohnisko. Z hlediska anotace zůstává výjimkou ze závazného pravidla a řadí se vpravo od svého řídicího uzlu.

Příklady:

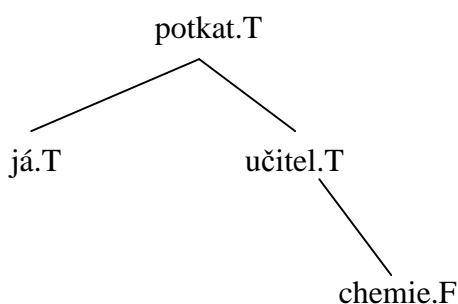
(6) *Novináři.T měli.F příležitost.F ocenit.F tento.T jinak.F skvělý.F vůz.T.*



(7) (*Mám rád červené tulipány, ale*) v Holandsku.C jsem viděl.F i.F (tulipány.T) modré.F.



(8) (*Kterého učitele jsi potkal?*) Potkal.T jsem učitele.T chemie.F.



Specifickým případem kvaziohniska jsou výrazy, které zahrnují názvy jednotek či jiná substantiva řídicí číslovky. Jsou to typická spojení, kdy je řídicí uzel známý z kontextu (jednotky patří mezi obecně sdílené informace) a důležitou novou informací je číslovka. Je ovšem nutné si uvědomit, že o kvaziohnisko se jedná pouze v případě, kdy je číslovka ve vlastním ohnisku – v ostatních případech se kontextově zapojené názvy jednotek řadí doleva od svého řídicího uzlu.

3.2. Pořadí v základu

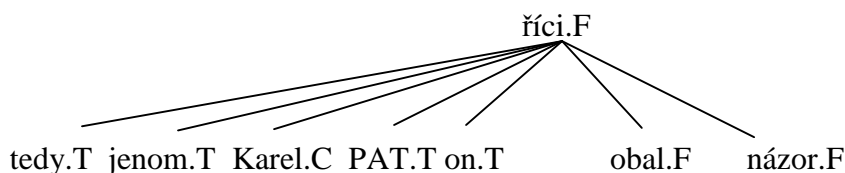
Pořadí uzlů v kontextově zapojené části věty nebylo zásadně přeuspořádáno, ale spíše dopracováno a zpodrobněno tak, jak se objevovaly další typy uzlů, které se v základu mohou vyskytnout. Těmi byly vokativní oslovení a uzly typu PREC a ATT, které byly umístěny úplně vlevo, protože se neúčastní větné struktury (rozvíjejí větu jako celek). Za ně byl umístěn rematizátor uvozující kontrastivní téma, protože rematizátor stojí vždy vlevo od uzlu, který má v dosahu; za ně pak kontrastivní téma, prvek, který je vždy centrem tématu, dále ostatní vyjádřené tematické uzly, jejichž pořadí bylo ponecháno v souladu s povrchoým především proto, aby mohly být sledovány případné tendence v rozmístění kontextově zapojených uzlů a souvislosti tohoto rozmístění se stupnicí výpovědní dynamičnosti. Následují ve shodě s předchozím manuálem uzly

elidované, slabé tvary zájmen a kulisy. Jako nejpravější část základu je uvedeno kontextově zapojené sloveso.

Důležité je zde i terminologické dopracování. Přestáváme používat termín „vlastní téma“, který sloužil především k označení centrálního prvku ve skupině kontextově zapojených výrazů, protože centrálnost v rámci základu je velmi obtížně určitelná, není-li signalizována kontrastem. Sledováním povrchového pořadí bude možné zjistit, nakolik právě toto pořadí určuje výpovědní dynamičnost v rámci základu. Opouštíme také termín „příklonky“, protože „druhá pozice“ v syntaktické struktuře je charakteristická i pro jiné výrazy, než jsou výrazy foneticky definované jako příklonky. Výrazy, které se vyskytují na druhé pozici ve větě, nemají v tektogramatickém stromě vlastní pozici také proto, že považujeme za vhodné řadit je podle významu.

Příklad:

(9) *Jenom.T Karel.C mu.T tedy.T včera.T (PAT.T) řekl.T bez obalu.F svůj názor.F.*



3.3. Pořadí v nominální skupině

Pořadí v nominální skupině bylo v původní verzi manuálu zpracováno poměrně vágně, kontextově nezapojené přívlastky (kterých je většina) měly být řazeny „podle relativní těsnosti“, což bylo příčinou velkého procenta neshod mezi anotacemi jednotlivých anotátorů, zvláště v případech, kdy se nezapojené přívlastky vyskytovaly v povrchovém pořadí vpravo i vlevo od substantiva. Proto byly v nominální skupině vysledovány typy přívlastků, které se jeví jako různé stupně těsnosti spojení s řídicím substantivem. Jako nejtěsnější vnímáme spojení řídicího uzlu a atributu ve víceslovném terminologickém spojení, názvu apod. Důležitá je i gramatická podoba atributu – nejtěsněji působí jmenovací nominativ, pak genitivní přívlastky, poměrně volně působí adjektivní atributy a ještě volněji předložkové pády. U přívlastkových vedlejších vět záleží na tom, zda je daná klauze restriktivní nebo nerestriktivní – nerestriktivní vztahná klauze, „volný přívlastek“, je pochopitelně nejvolnější a tedy nejdynamičtější v substantivní skupině. V jednotlivých skupinách je opět sledován vztah mezi povrchovým slovosledem a výpovědní dynamičností – preponované přívlastky se řadí v opačném pořadí než je povrchové, postponované ho zpravidla dodržují – zdá se platit, že čím vyšší je výpovědní dynamičnost atributu, tím dále je na povrchu od řídicího substantiva. K tomuto řazení jsme se tedy v nové verzi manuálu přiklonili - pochopitelně kromě případů, které by z této tendence svým významem výrazně vybočily, neboť významové hledisko zůstává nejdůležitějším kritériem.

3.4. Neprojektivní konstrukce

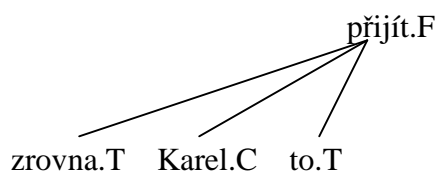
U vět, ve kterých některý uzel v povrchovém pořadí porušuje pravidlo projektivity tektogramatického stromu, bylo třeba nově zpracovat anotaci neprojektivních uzlů v případech frazeologických spojení a některých složených slovesných predikátů, u kterých starší verze manuálu počítala s tím, že budou v tektogramatické struktuře reprezentovány jediným uzlem. Rozhodnout, kterých spojení se to bude týkat, je ale možné až po dokončení valenčního slovníku substantiv, proto bylo nutné vypracovat provizorní řešení pro případy, kdy je struktura zdánlivě neprojektivní a tato neprojektivita není motivována užitím kontrastivního tématu. Protože zprojektivněné uzly, které jsou hodnoceny jako kontextově zapojené, nenesou na rozdíl od těch kontrastivních informaci o tom, že byly přesunuty z neprojektivní polohy, bylo rozhodnuto, že kvůli zachování informace o původním umístění uzlu bude v těchto konstrukcích prozatím ponechána neprojektivní hrana, což ovšem neznamená rezignaci na pravidlo projektivity tektogramatické struktury, pouze možnost pro další výzkum tohoto jevu.

3.5. Rematizátory

Jedním z důležitých zjištění v posledním období bylo, že úlohou rematizátoru, který se vyskytuje v základu, je nejspíš signalizace kontrastivního tématu, jako v Příkladu (10). Pokud rematizátor signalizuje kontrastivitu tématu, v jeho dosahu je pouze jeho pravý sesterský uzel (popř. koordinovaná skupina). Protože takto postavený rematizátor zapojuje kontrastivní téma do kontextu (stojí většinou na začátku klauze), značkujeme ho jako kontextově zapojený a zůstává vlevo od uzlu v jeho dosahu jako ostatní rematizátory.

Příklad:

(10) *Zrovna.T Karel.C na to.T přijít.F.*

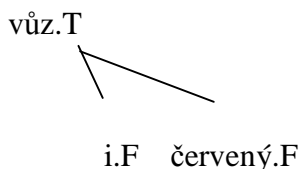


Problematickým se ukázal dosah rematizátoru v nominální skupině. U spojení jako *i červený vůz* je dosah rematizátoru jednoznačný, pokud je řídicí substantivum kontextově zapojené – pak podle polohy ve větě značkujeme jako v příkladě (11a) nebo (11b). Problém ovšem nastává, není-li rematizátor kontextově zapojený. Protože shodný adjektivní přívlastek musí stát z gramatických důvodů před řídicím substantivem, může mít rematizátor stojící před ním v dosahu buď celou skupinu, nebo pouze toto adjektivum. Pak jsou u těchto skupin možné dvě interpretace – jako v příkladech (12) a (13). Jediným

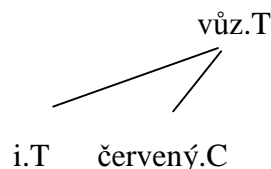
spolehlivým ukazatelem je zde ovšem poloha intonačního centra, kterou v psané podobě textu nemůžeme zjistit. Proto anotujeme všechny tyto skupiny jako varianty (12a) a (12b), které jsou bezpříznakové. Ke zjištění, v jaké míře a jestli vůbec se v textech vyskytují příznakové varianty (13a) a (13b), bude třeba zkoumat mluvené texty.

Příklady:

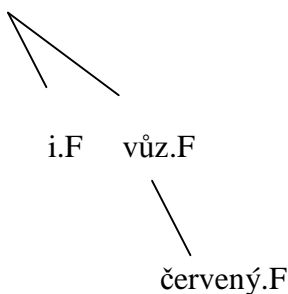
(11a)



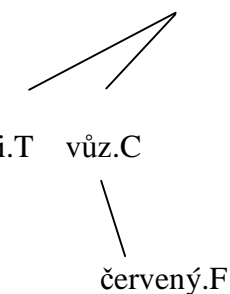
(11b)



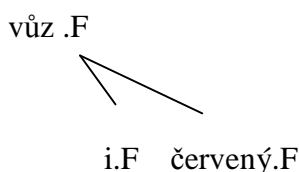
(12a)



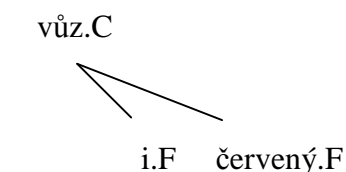
(12b)



(13a)



(13b)



3.6. Aktuální členění otázky

Položkou, která dosud v manuálu zcela chyběla, bylo značkování otázky. Vzhledem k rostoucímu zastoupení žánru rozhovoru, v němž jsou otázky velmi frekventované, v anotovaném materiálu bylo třeba navrhnout nějaké řešení.

Otázka zjišťovací může být v češtině tvořena dvěma způsoby - intonací nebo inverzním slovosledem. Pokud je otázka tvořena pomocí intonace, je zachován slovosled oznamovací věty i její aktuální členění - před slovesem jsou zpravidla zapojené uzly, za slovesem nezapojené a zcela vpravo vlastní ohnisko věty (viz př. (15)). I zde se může uplatňovat tendence slovesa stát na druhé pozici ve větě (viz př. (15e)). Pokud je otázka tvořena inverzním slovosledem, vypadá situace poněkud složitěji - sloveso se na povrchu posouvá na začátek věty, vlastní ohnisko se (v bezpříznakových případech) řadí na její

konec (viz př. (14)). Protože sloveso stojí na začátku, není ovšem signalizován předěl mezi uzly zapojenými a těmi nezapojenými, které leží mezi slovesem a vlastním ohniskem. Jediným vodičkem kromě významu zde může být porušování systémového pořadí (viz př. (16)). V tektogramatickém stromě se tedy zjišťovací otázka anotuje tak, že sloveso (i když leží na povrchu na začátku) se řadí jako nejméně dynamická část ohniska, vlastní ohnisko úplně doprava, mezi ně nezapojené uzly v pořadí shodném s povrchovým. Zapojené uzly podléhají stejným zásadám pro pořadí v základu jako v oznamovací větě.

Doplňovací otázka je tvořena tázacím výrazem. Tento výraz je tím, na co se produktor otázky ptá, proto bývá vlastním ohniskem, ale nemusí jím být. V příkladech jako (17) je vlastním ohniskem výraz na konci věty, který také nese intonační centrum. Tázací výraz je pak anotován jako nejméně dynamická část ohniska a řadí se vpravo za sloveso. Předěl mezi základem a ohniskem signalizuje sloveso, stejně jako v oznamovací větě, v doplňovací otázce však stojí sloveso poměrně často na druhé pozici ve větě hned za tázacím výrazem. Mezi tázacím výrazem a slovesem může na povrchu stát kontrastivní uzel (18).

I v otázce ovšem existuje subjektivní pořadí (viz př. (19)). Při jeho anotaci se postupuje opět tak, že je ve stromě přeřazeno stejně jako bezpříznakové případy.

Příklady:

(14a) *Pojedeš.F letos.T v létě.T NA HORY.F?*

(14b) *Pojedeš.F na hory.T letos.F V LÉTĚ.F?*

(14c) *Pojedeš.F letos.T na hory.F V LÉTĚ.F?*

(15a) *Letos.T v létě.T pojedeš.T/F NA HORY.F?*

(15b) *Na hory.C pojedeš.F letos.F V LÉTĚ.F?*

(15c) *Na hory.C letos.T pojedeš.T V LÉTĚ.F?*

(15d) *Na hory.T letos.T v létě.T POJEDEŠ.F?*

(15e) *Letos.C pojedeš.F v létě.T NA HORY.F?*

(16) *Pojede.F Jirka.C letos.T v létě.T NA HORY.F?*

(17) *Kam.F pojedeš.T NA DOVOLENOU.F?*

(18) *Kam.F Jirka.C pojede.F NA DOVOLENOU.F?*

(19a) *NA DOVOLENOU.F Jirka.T pojede.T?*

(19b) *Na dovolenou.C pojedeš.T KAM.F?*

3.7. Drobnější změny

Kromě dopracování větších celků byla v manuálu provedena řada drobnějších úprav. Byl rozšířen příkladový aparát, pro větší přehlednost bylo změněno členění manuálu a jednotlivé pokyny byly hierarchizovány, bylo doplněno vysvětlení základních pojmů zpracování titulků, nevětných konstrukcí, parentezí, vokativů, pomocná kritéria pro anotaci kontrastivního tématu a byly rozvedeny a zpodrobněny některé oddíly, které byly zpracovány už v původním manuálu. Pro zajímavost celý manuál narostl z původního rozsahu na více než dvojnásobek.

4. Výhledy

4.1. Kontrastivní téma

Přiřazení hodnoty *c* je v manuálu prozatím formulováno dosti vágně, především proto, že není ještě dostatečně prozkoumáno, kde všude se může kontrast vyskytnout a které faktory mají vliv na jeho použití. V současné době víme, že

- na výskyt kontrastivního tématu má rozhodující vliv tématická struktura textu, kontrastivní téma se vyskytuje v jednotlivých oddílech výčtových pasáží, na začátcích odstavců apod.,
- některé typy uzlů mají silnější sklony být využity jako kontrastivní téma (např. aktanty spíše než volná doplnění),
- kontrastivní uzel může být ve struktuře zapuštěný libovolně hluboko,
- kontrastivní uzel nese v mluvené podobě fakultativní kontrastivní přízvuk (viz Veselá et al. 2003),
- kontrastivní téma může být signalizováno rematizátorem,
- věta může mít více než jedno kontrastivní téma,
- kontrastivní téma se zpravidla vyskytuje v iniciální pozici ve větě,
- kontrast může být vysvětlen tím, že označuje výběr z množiny alternativ a může být různě silný podle podoby této množiny.

V budoucnu je třeba vyhodnotit příklady, které odporují některým základním tendencím (např. věty s rematizátorem uprostřed základu) a zjistit, které faktory působí na které vlastnosti centra základu a které způsobují jeho obsazení kontrastivním uzlem. Je nutné především najít kritérium pro rozlišení kontrastivního tématu od kontextově zapojeného uzlu v případech, kde jsme zatím odkázáni na přízvuk.

4.2. Zpracování složených predikátů a klasifikace povrchových neprojektivit

Velké množství povrchově neprojektivních konstrukcí vzniká ve stromech, v nichž se predikát skládá z více jednotek a některé participanty závisí na jeho závislé části; jedná se buď o tzv. složené predikáty – konstrukce složené z významově vyprázdněného slovesa a deverbativního substantiva, nebo o konstrukce s modálním a fázovým významem složené ze slovesa a infinitivu, nebo o některá částečně ustálená spojení. Slovesná konstrukce se zde z hlediska aktuálního členění chová jako jediná jednotka a kontextově zapojené uzly na ní závislé stojí bezpříznakově na začátku věty – v takovém případě ovšem strom přestává být projektivní a jedná se o neprojektivitu způsobenou formálně. Řešením pro značkování aktuálního členění by bylo, kdyby tyto kontextově

zapojené uzly závisely na řídicím slovese a ne na závislém deverbativu či infinitivu. Proti tomu však hovoří v mnoha případech argumenty na straně značkování syntaktických funkcí. Konečné řešení závisí především na podrobné klasifikaci jednotlivých případů, především odstupňování těsnosti jednotlivých spojení.

Není také dosud zcela jasné, kdy je neprojektivita motivována použitím kontrastivního tématu a kdy jím být motivována nemusí. Proto je třeba zmapovat různé podoby neprojektivních konstrukcí (nejen ty, v nichž je uzel předsunut neprojektivně doleva) a popsát, jak s nimi na tektogramatické rovině zacházet.

4.3. Rematizátory v nominální skupině

Jak bylo řečeno výše, problematická je anotace rematizátorů v nominální skupině. Tento problém však sahá ještě o něco dále – u vazeb s genitivními přívlasky často nelze vložit rematizátor dovnitř nominální skupiny. Zvláště u skupin rozvitějších lze pak velmi obtížně určit dosah rematizátoru a jeho pozici v hloubkové struktuře. Proto bude třeba otestovat chování podobných skupin a vysledovat případné tendence v chování rematizátorů.

Problémem jsou i rematizátory vyskytující se uprostřed základu.

4.4. Vertikální členění textu

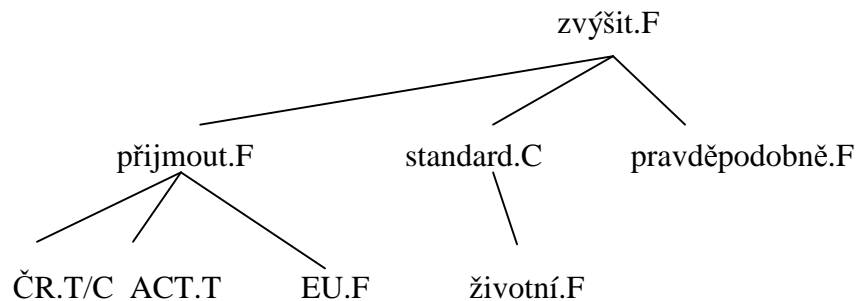
I když je v manuálu nově zpracováno anotování titulků, zůstává v oblasti vertikálního členění textu ještě dost problémů. Jde o jevy jako podtitulky, popisky obrázků, poznámky pod čarou a podobně. Klíčovým úkolem zde bude specifikovat úlohu vertikálního členění při utváření informační struktury textu.

4.5. Podřadné souvětí

Při anotacích bylo zjištěno, že některá podřadná souvětí, zejména příčinného typu, se z hlediska aktuálního členění chovají jako souřadná. Informace uváděná v závislé klauzi (zejména stojí-li tato vlevo od klauze řídicí) uvádí děj z hlediska kontextové zapojenosti natolik nezávislý na ději řídicí klauze, že se dá jen velice těžko interpretovat jako součást základu nebo ohniska řídicí klauze. Problémem je především najít popsatelnou hranici, kterou povedeme mezi případy, kdy se anotuje závislá klauze jako součást řídicí klauze, a těmi, kde se anotuje jako samostatné sdělení. Je nutné rovněž podrobit zkoumání tendence u jednotlivých typů těchto souvětí (příčinných, podmínkových, přípustkových atd.).

Příklad:

(20) *Pokud bude ČR.C přijata.F do EU.F, životní.F standard.C se pravděpodobně.F zvýší.F.*



Bylo též zjištěno, že některé participanty se chovají jinak z hlediska pořadí, když jsou vyjádřeny větňě, než když jsou vyjádřeny člensky. I zde je třeba probrat nalezené příklady a popsat zjištěné tendence.

5. Technické řešení v Pražském závislostním korpusu

5.1. Závislostní stromy

K zachycení syntaktické struktury v Pražském závislostním korpusu se používají závislostní stromy s lineárním uspořádáním na uzlech. To znamená, že na uzlech existují dvě uspořádání: částečné uspořádání zachycující syntaktickou závislost uzlů a úplné uspořádání zachycující pořadí uzlů „zleva doprava“. V následujícím budeme mluvit o závislosti a podřízenosti (což je tranzitivní uzávěr relace závislosti) v případě částečného uspořádání zachycujícího syntaktickou strukturu závislostního stromu, pro vyjadřování o lineárním uspořádání budeme používat opisy „vlevo“, „vpravo“, „mezi“ apod. (Lineární uspořádání určuje pro každé dva uzly jejich vzájemné pořadí v tomto uspořádání.) Nadále předpokládáme, že na každém závislostním stromu je definováno lineární uspořádání na uzlech.

5.1.1. Podmínka projektivity závislostních stromů

Z hlediska teoretického rámce užitého při návrhu tektogramatické roviny Pražského závislostního korpusu (Funkčního generativního popisu) je zásadní podmínkou kladenou na tvar tektogramatických závislostních stromů tzv. podmínka projektivity.

Podmínku projektivity lze definovat různými způsoby, které jsou ekvivalentní, aplikujeme-li je na celý závislostní strom s lineárním uspořádáním, avšak mohou se lišit při jejich aplikování na vlastní podstromy nějakého závislostního stromu. My zde uvedeme dvě možné definice pojmu projektivita ilustrující zmíněnou situaci. Protože nezavádíme žádný formalismus pro popis závislostních stromů s lineárním uspořádáním na uzlech, bude následující výklad (snad ne nepřiměřeně) neformální.

Definice 1. (lokální projektivita) Podstrom P závislostního stromu S je projektivní (splňuje podmínku projektivity), pokud pro každý jeho vnitřní uzel a (tj. uzel podstromu P různý od kořene podstromu P) platí, že relativní pořadí všech uzlů b podstromu uzlu a je v lineárním uspořádání vůči rodiči a sourozencům uzlu a stejné.

Definice 2. (totální projektivita, pomocí pokrytí) Podstrom P závislostního stromu S je projektivní (splňuje podmínku projektivity), pokud pro každý podstrom T podstromu P platí, že uzly podstromu T pokrývají souvislý podúsek uspořádaného seznamu uzlů celého závislostního stromu S .

Přiblížme si obě definice názorně: První definice říká, že když shora dolů procházíme podstrom, jehož projektivitu ověřujeme, a závislostní hrana nás „vede“ z nějakého uzlu x do nějakého uzlu a , potom se už žádný uzel b v podstromu uzlu a nesmí dostat v lineárním uspořádání doleva ani doprava přes uzel x ani sourozence uzlu a .

Druhá definice pro splnění projektivity libovolného podstromu P požaduje, aby každý jeho podstrom T odpovídal souvislému podúseku věty, čili jinými slovy zabraňuje tomu, aby se mezi uzly podstromu T dostal „cizí“ uzel nepatřící do tohoto podstromu (raději připomeňme, že „mezi“ se zde vztahuje k lineárnímu uspořádání na uzlech).

Abychom rozptýlili případné čtenářovy pochybnosti, první definice je slabší než druhá, vlastní podstrom nějakého závislostního stromu, který je projektivní podle druhé definice, je vždy také projektivní podle první definice, ale ne nutně naopak.

Je snadné nahlédnout, že pokud Definici 2. oslabíme tak, že budeme požadovat pokrytí souvislého podúseku v lineárním uspořádání zúženém na podstrom P , dostaneme definici ekvivalentní s Definicí 1. Odsud zřejmě plyne ekvivalence Definic 1. a 2. pro celé závislostní stromy.

5.1.2. Algoritmus pro zprojektivnění/kontrolu projektivity

Protože se při anotaci tektogramatické roviny Pražského závislostního korpusu postupuje v podstatě tak, že se jako první přiblížení tektogramatické struktury věty přebírá závislostní struktura i lineární uspořádání z analytické roviny (v zájmu zjednodušení opomíjíme některé součásti anotačního schématu, které však situaci z hlediska projektivity závislostních stromů nijak nekomplikují), bylo nezbytné navrhnout algoritmus, který by mohl sloužit nejen ke kontrole projektivity daného závislostního stromu, ale mohl být také použit k převedení libovolného závislostního stromu na strom projektivní.

Níže uvedený algoritmus odpovídá Definici 1. uvedené v předchozím oddílu; jeho myšlenka je prostá a byla už naznačena při neformálním přiblížení významu definice. Postupujeme tak, jako bychom závislostní strom postupně „rozčesávali“ – postupujeme od „konečků“ směrem nahoru k uzlu, jehož podstrom chceme zprojektivnit, a když zpracováváme nějaký uzel, rozdělíme si uzly na něm závislé (jeho „děti“) pomyslným hřebenem tak, že mezi každé dvě děti zapadne jeden zub hřebenu, takže se uzly podřízené zpracovávanému uzlu postupně zdola „rozšmodrchávají“.

Algoritmus uvádíme v jednoduchém imperativním pseudokódu:

```
procedure Zprojektivni(uzel) {
    deti := uzel.deti;
    /* seznam uzlů závislých na zpracovávaném uzlu */
    foreach dite in deti do Zprojektivni(dite);
    /* rekurzivně zprojektivni podstromy všech dětí zpracovávaného uzlu */
    preusporadej_strom(uzel);
    /* přeuspořádej strom tak, aby podstromy dětí zpracovávaného uzlu byly
    v lineárním uspořádání bezprostředně kolem zpracovávaného uzlu, a to se
    zachováním vzájemného pořadí zpracovávaného uzlu a jeho dětí */
}
```

Popišme si algoritmus podrobněji: Algoritmus je rekurzivní, aplikuje se při průchodu do hloubky na každý uzel zpracovávaného podstromu – procedura dostává právě zpracovávaný uzel jako parametr. Procedura se napřed rekurzivně zavolá na všechny děti zpracovávaného uzlu, a tím zprojektivní všechny odpovídající podstromy (pokud na zpracovávaném uzlu nezávisí žádný uzel, rekurze končí). Zásadní je v algoritmu podprocedura zajišťující přeuspořádání podstromu zpracovávaného uzlu. Ta musí zajistit přeuspořádání podstromu zpracovávaného uzlu tak, aby vzájemné pořadí zpracovávaného uzlu a podstromů jeho dětí odpovídalo vzájemnému pořadí zpracovávaného uzlu a jeho dětí a podstromy přitom byly v lineárním uspořádání bezprostředně kolem zpracovávaného uzlu. (Jinými slovy, podstromy dětí přesuneme co nejbližší ke zpracovávanému uzlu, a to bez porušení vzájemného pořadí zpracovávaného uzlu a jeho dětí.)

Algoritmus je možné modifikovat tak, aby mohl být použit jak ke zprojektivnění daného podstromu, tak ke kontrole projektivity jak podle Definice 1., tak podle Definice 2. z předchozího oddílu.

Myšlenka použití algoritmu ke kontrole projektivity je snadná: vytvoříme kopii (pod)stromu, jehož projektivitu chceme ověřit, tuto kopii zprojektivníme a výsledek porovnáme co do pořadí s původním (pod)stromem.

Chceme-li ověřit projektivitu podstromu nějakého závislostního stromu podle Definice 1., vytvoříme kopii pouze daného podstromu, tuto kopii zprojektivníme a porovnáme relativní pořadí uzlů daného podstromu s jeho zprojektivněnou kopií. Pokud se pořadí v podstromu a jeho zprojektivněné kopii liší, podstrom není projektivní.

Pro ověření projektivity podle Definice 2. je navíc oproti postupu v předešlém odstavci potřeba ověřit, že podstrom v celém závislostním stromu pokrývá souvislý úsek vzhledem k lineárnímu uspořádání na uzlech. Pokud tomu tak není, strom není projektivní podle Definice 2.

Složitost algoritmu je závislá na datové reprezentaci a použití pomocných datových struktur. Převedením rekurze na cyklus a použitím pomocného pole jsme při implementaci algoritmu v editoru stromových struktur TrEd dosáhli lineární časové i prostorové složitosti vzhledem k počtu uzlů zpracovávaného podstromu.

5.2. Programové nástroje pro modifikaci stromové struktury

Při anotování Pražského závislostního korpusu se používá editor stromových struktur TrEd (Hajič et al. 2001, Pajas 2003) naprogramovaný Petrem Pajasem. Editor je naprogramovaný v jazyce Perl a je navržený tak, že umožňuje doplňování modulů rozšiřujících jeho možnosti. Nejprve stručně nastíníme datovou reprezentaci stromových struktur v TrEd-u, v další části popíšeme modul pro anotování TFA.

5.2.1. Datová reprezentace závislostních stromů v editoru stromů TrEd

Pro úplnost stručně naznačme datovou reprezentaci používanou v nástroji pro anotování Pražského závislostního korpusu TrEd. Uzly závislostního stromu jsou reprezentovány tabulkou s položkami reprezentujícími jak lingvistické informace (lemma, funktor, gramatémy, různé další atributy), tak informace spíše technického rázu (zejména informace týkající se struktury závislostního stromu, lineárního uspořádání na uzlech a jiné).

K zachycení stromové struktury slouží následující čtyři položky v tabulce: `parent`, `firstson`, `lbrother`, `rbrother`. První dvě položky obsahují ukazatele na rodiče a „prvního“ syna (v alternativní terminologii dceru) – slouží tedy k zachycení vertikální struktury závislostního stromu, druhé dvě zprostředkovávají přístup k „sousedním“ uzlům, tj. uzlům závislým na témž rodiči – slouží tedy k zachycení horizontální struktury závislostního stromu. (Uvozovky jsou v předchozí větě proto, že reprezentace závislostní, tj. stromové, struktury je technicky nezávislá na lineárním uspořádání uzlů, takže např. syn, na nějž ukazuje položka `firstson`, nemusí být v lineárním uspořádání na uzlech první.)

Lineární uspořádání na uzlech je zachyceno zvláštní položkou, v případě tektogramatických stromových struktur se jedná o atribut `dord`.

5.2.2. Makra pro editor stromů TrEd

Pro anotování TFA na tektogramatické rovině Pražského závislostního korpusu jsme vyvinuli modul obsahující vhodné funkce rozšiřující možnosti editoru TrEd, tzv. makra. Makra jsou sloučená do tzv. anotačního módu TFA. Dají se zhruba rozdělit na makra modifikující pořadí uzlů (podstromů) v závislostním stromu a na makra pro přiřazování hodnot atributu `tfa` jednotlivým uzlům stromu. Uvádíme pouze makra přístupná anotátorům při práci s editorem v módu TFA, pomocné funkce neuvádíme (každému níže uvedenému makru je pro urychlení práce také přiřazena zde neuváděná klávesová zkratka).

5.2.2.1. Makra pro přiřazování hodnoty atributu `tfa`

Základní makra slouží k přiřazení hodnot do atributu `tfa`.

<code>tfa_focus</code>	Do atributu <code>tfa</code> aktivního uzlu přiřadí hodnotu F.
<code>tfa_topic</code>	Do atributu <code>tfa</code> aktivního uzlu přiřadí hodnotu T.
<code>tfa_C</code>	Do atributu <code>tfa</code> aktivního uzlu přiřadí hodnotu C.
<code>tfa_NA</code>	Do atributu <code>tfa</code> aktivního uzlu přiřadí hodnotu NA.
<code>tfa_qm</code>	Do atributu <code>tfa</code> aktivního uzlu přiřadí hodnotu ???.

Další skupinu tvoří makra automaticky přiřazující hodnoty do atributu `tfa`.

<code>PreSetTFACurrentST</code>	Na základě heuristických lingvistických pravidel přiřadí hodnoty do atributů <code>tfa</code> uzlů z podstromu aktivního uzlu.
<code>PreSetTFATree</code>	Totéž, ale pro celý strom, v němž se nachází aktivní uzel.
<code>PreSetTFAAllTrees</code>	Totéž, ale pro všechny stromy v souboru, v němž se nachází aktivní uzel.

5.2.2.2. Makra pro modifikaci pořadí uzlů/podstromů

Makra modifikují pouze uspořádání uzlů, nemění závislostní strukturu stromu. Uvádíme přehled maker s doplňujícím komentářem.

Nejjednodušší makra umožňují přesun jednotlivých uzlů v lineárním uspořádání uzlů.

<code>ShiftLeft</code>	Posune aktivní uzel doleva (tj. „ob uzel“).
<code>ShiftRight</code>	Posune aktivní uzel doprava.

Ostatní makra pracují s podstromy, a aby jejich funkčnost byla vůbec rozumně definovatelná (jak si čtenář laskavě rozmyslí), musí podstrom, na který se aplikují, splňovat podmínku projektivity.

<code>ProjectivizeCurrentSubTree</code>	Zkontroluje projektivitu podstromu aktivního uzlu; pokud podstrom není projektivní, nabídne možnost jeho zprojektivnění.
<code>ProjectivizeTree</code>	Totéž, ale pro celý strom, v němž se nachází aktivní uzel.

Makra přesouvající podstromy používají kontrolu projektivity, takže pokud by se měla aplikovat na neprojektivní podstromy, vyžadují po anotátorovi souhlas s jejich zprojektivněním.

<code>ShiftSTLeft</code>	Posune podstrom aktivního uzlu doleva o jeden uzel (podstrom aktivního uzlu se jako celek prohodí s prvním uzlem nepatřícím do tohoto podstromu, který se nachází od něj vlevo).
--------------------------	--

ShiftSTRight	Posune podstrom aktivního uzlu doprava o jeden uzel.
ShiftSTOverSTLeft	Prohodí podstrom aktivního uzlu s podstromem levého sourozence aktivního uzlu.
ShiftSTOverSTRight	Prohodí podstrom aktivního uzlu s podstromem pravého sourozence aktivního uzlu.
OrderSTByTFA	Přeuspořádá podstromy dětí aktivního uzlu podle hodnoty jejich atributů $t f a$ tak, aby jejich pořadí odpovídalo pořadí C, T, F, a relativní pořadí v jednotlivých skupinách zůstalo zachováno.

Literatura

- DOKULIL, Miloš; DANEŠ, František. K tzv. mluvnické a významové stavbě věty. In *O vědeckém poznání soudobých jazyků*. Praha: Nakladatelství ČSAV, 1958. S. 231-246.
- HAIJČ, Jan; HAIJČOVÁ, Eva; ROSEN, Alexandr. Formal Representations of Language Structures. TELRI Newsletter, June 1996, no 3, s. 12-19.
- HAIJČ, Jan; HAIJČOVÁ, Eva. Syntactic Tagging in the Prague Tree Bank. In *Proceedings of the Second European Seminar "Language Applications for a Multilingual Europe"*. Ed. by Marcinkeviciene, Ruta; Volz, Norbert. Kaunas; Manheim: LDS;VDU, 1997. Pp. 55-68.
- HAIJČ, Jan; HAIJČOVÁ, Eva; PANEVOVÁ, Jarmila; SGALL, Petr. *Specifikace stromových struktur na tektogramatické úrovni značkování a základní kroky překladač z analytické na tektogramatickou úroveň*. Technická zpráva projektu GAČR 405-96-0198 "Formální reprezentace jazykových struktur". Praha: Karlova Univerzita, 1998.
- HAIJČ, Jan; HAIJČOVÁ, Eva; PANEVOVÁ, Jarmila; SGALL, Petr. Syntax v českém národním korpusu. Slovo a slovesnost, roč. 59, č. , 1998. S. 168-177.
- HAIJČ, Jan. Building a syntactically annotated corpus: The Prague Dependency Treebank. In *Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová*. Ed. by Hajičová, Eva. Praha: Karolinum, 1998. Pp. 106-132.
- HAIJČ, Jan; VIDOVÁ-HLADKÁ, Barbora; PAJAS, Petr. The Prague Dependency Treebank: Annotation Structure and Support. In *Proceeding of the IRCS Workshop on Linguistic Databases*. Ed. Bird, Steven; Buneman, Peter; Liberman, Mark. Philadelphia: University of Pennsylvania, 2001. Pp. 105-114.
- HAIJČ, Jan; VIDOVÁ-HLADKÁ, Barbora; HAIJČOVÁ, Eva; SGALL, Petr; PAJAS, Petr; ŘEZNÍČKOVÁ, Veronika; HOLUB, Martin (2001). The Current Status of the Prague Dependency Treebank. In *Text, Speech and Dialogue, 4th International Conference, TSD 2001, Zelezná Ruda, Czech Republic, September 11-13, 2001, Proceedings*. Eds. Matoušek, Václav; Mautner, Pavel; Mouček, Rudolf; Taušer; Karel. Berlin; Heidelberg: Springer-Verlag, 2001. LNAI 2166, ISBN 3-540-42557-8. Pp. 11-20.
- HAIJČ, Jan; VIDOVÁ-HLADKÁ, Barbora; PANEVOVÁ, Jarmila; HAIJČOVÁ, Eva; SGALL, Petr; PAJAS, Petr (2001): Prague Dependency Treebank 1.0 (Final Production Label). Published by Linguistic Data Consortium on *CDROM*, LDC Catalog no.: LDC2001T10., ISBN 1-58563-212-0.
- HAIJČOVÁ, Eva. *Issues of sentence structure and discourse patterns*. Praha: Karlova Univerzita, 1993.

- HAIČOVÁ, Eva; SGALL, Petr; SKOUMALOVÁ, Hana. An automatic procedure for topic-focus identification. *Computational Linguistics*, vol. 21, 1995. Pp. 81-94.
- HAIČOVÁ, Eva. Prague Dependency Treebank: From analytic to tectogrammatical annotations. *Proceedings of the 1st International Workshop on Text, Speech, Dialogue*, Brno, Czech Republic, September 1998. Eds. Sojka, Petr; Matoušek, Václav; Pala, Karel; Kopeček, Ivan. Brno: Masarykova univerzita, 1998. ISBN 80-210-1899-2. Pp. 45-50.
- HAIČOVÁ, Eva, PANEVOVÁ, Jarmila; SGALL, Petr. Language Resources Need Annotations To Make Them Really Reusable: The Prague Dependency Treebank. In *Proceedings of the First International Conference on Language Resources & Evaluation, Granada, Spain, 28-30 May 1998*. Granada: University of Granada, 1998. Pp. 713-718.
- HAIČOVÁ, Eva (2000): How Many Topics/Foci. In *Linguistics and Language Studies; Exploring language from different perspectives* Eds. I. Kovačič, M. Milojević-Sheppard, S. Orel-Kos, J. Orešnik. Ljubljana: Filozofska fakulteta Univerze v Ljubljani, 2000. ISBN 961-227-071-6. Pp. 9-20.
- HAIČOVÁ, Eva. Mysteries of Order. In *Člověk a jeho jazyk; 1. Jazyk jako fenomén kultury (na počest profesora Jána Horeckého)*. Ed. Kl. Buzássyová. Bratislava: Veda, vydavateľstvo Slovenskej akadémie vied, 2000. ISBN 80-224-0641-4. S. 260-268.
- HAIČOVÁ, Eva. Teorie optimality a aktuální členění věty. *Slovo a slovesnost*, roč. 61, č. 3, 2000. S. 161-169.
- HAIČOVÁ, Eva. Item Ordering in the Sentence. In *Proceedings of LP'98. Item Order in Language and Speech Columbus, The Ohio State University, September 15-20, 1998*. Eds. Fujimura, Osamu; Joseph, Brian D.; Palek, Bohumil. Praha: Karolinum, 1999. ISBN 80-246-0016-1. Pp. 361-371.
- HAIČOVÁ, Eva; SGALL, Petr; BURÁŇOVÁ, Eva. Tagging of a very large corpora: Topic-Focus Articulation. In *The 18th International Conference on Computational Linguistics Coling 2000 in Europe: Proceedings of the Conference Universitat Des Saarlandes Saarbrücken, Germany 31 July - 4 August 2000*. Saarbrücken: Universität Des Saarlandes, 2000. ISBN 1-55860-717-X. Pp. 139-144.
- HAIČOVÁ, Eva. Possibilities and Limits of Optimality in Topic-Focus Articulation. In *Current issues in formal Slavic linguistics*. Eds. Zybatow, G.; Junghanns, U.; Mehlhorn, G.; Szucsich L. Frankfurt/Main: Peter Lang, 2001. Pp. 385-394.
- HAIČOVÁ, Eva; SGALL, Petr. Dependency, Coordination, and Projectivity. In *Slovo v tekste i v slovaru (sbornik statej k semidesjatiletiju akademika Ju. D. Apresjana)*. Eds. Iomdin, L.L.; Krysin, L.P. Moskva: Izd. Jazyki ruskij kul'tury. *Studia Philologica*, 2001. ISBN 5-7859-0199-4. Pp.456-466.
- HAIČOVÁ, Eva; SGALL, Petr. Topic-focus and salience. In *Association for Computational Linguistic, 39th Annual Meeting and 10th Conference of the*

- European Chapter, Proceedings of the Conference, July 9-11, 2001, Toulouse, France.* Toulouse: Morgan Kaufmann Publishers, 2001. Pp. 268-273.
- HAIČOVÁ, Eva; SGALL, Petr; HAVELKA, Jiří (2003): Discourse Semantics and the Salience of Referents. In *Journal of Slavic Linguistics*, pp. (submitted).
- HAIČOVÁ, Eva: Theoretical description of language as a basis of corpus annotation: The case of Prague Dependency Treebank. In *Prague Linguistic Circle Papers. Volume 4*. Eds. Hana, Jiří; Sgall, Petr; Hajičová, Eva; Hoskovec, Tomáš. Amsterdam: John Benjamins, 2002. Pp. 111-127.
- HAIČOVÁ, Eva; PAJAS, Petr; VESELÁ, Kateřina (2002): Corpus Annotation on the Tectogrammatical Layer: Summarizing the First Stages of Evaluations. *Prague Bulletin of Mathematical Linguistics*, no. 77, 2002. Pp. 5-18.
- HAIČOVÁ, Eva: Syntactic theory and corpus annotation need each other. In *Zbornik povzetkov. XIII. mednarodni slavistični kongres, 2. del*. Ed. Novak, France. Ljubljana: Mednarodni slavistični komite, 2003. Pp. 289.
- HAIČOVÁ, Eva: Topic-focus articulation in the Czech National Corpus. In *Language and function. To the memory of Jan Firbas*. Ed. Hladký, Josef. Amsterdam: John Benjamins, 2003. Pp. 185-194.
- HAIČOVÁ, Eva: Information structure and syntactic complexity. In *Investigations into formal Slavic linguistics. Contributions of the Fourth European Conference on Formal Description of Slavic Languages – FDSL IV Held at Potsdam University, November 28-30, 2001*. Ed. Kosta, Peter. New York: Peter Lang, 2003. Pp. 169-180.
- HAIČOVÁ, Eva (2003): Aspects of discourse structure. In *Natural language processing between linguistic inquiry and system engineering*. Eds. Menzel, Wolfgang; Vertan, Christina. Iasi: Editura Universitatii "Alexandru Ioan Cuza", 2003. Pp. 47-54.
- HAIČOVÁ, Eva (2003): Contextual boundness and discourse patterns. In *CIL 17 Proceedings*, pp.. Benjamins (in press).
- HAIČOVÁ, Eva; SGALL, Petr (2003): Information Structure, Translation and Discourse. In *Textologie und Translation : Jahrbuch Übersetzen und Dolmetschen 4/II*. Eds. Gerzymisch-Arbogast, Heidrun et al. Tübingen: Gunter Narr, 2003. Pp. 107-123.
- HAIČOVÁ, Eva; SGALL, Petr; Buráňová, Eva (2003): Topic-Focus Articulation and degrees of salience in the Prague Dependency Treebank. In *Formal Approaches to Function in Grammar. In honor of Eloise Jelinek, Arizona*, pp. 165--177. John Benjamins.
- HAIČOVÁ, Eva; SGALL, Petr; VESELÁ, Kateřina (2003): Information structure and contrastive topic. In *Formal Approaches to Slavic Linguistics #11: The Amherst Meeting 2002*. Eds. Browne, Wayles; Ji-Yung, Kim; Partee, Barbara H. and

- Rothstein, Robert A. Ann Arbor: Michigan Slavic Publications, 2003. Pp. 219-234.
- PAJAS, Petr (2003). Tree Editor TrEd, <http://ckl.mff.cuni.cz/pajas/tred/>
- PANEVOVÁ, Jarmila; BENEŠOVÁ, Eva; SGALL, Petr. *Čas a modalita v češtině*. Praha: AUC - Philologica Monographia 34, 1971.
- PANEVOVÁ, Jarmila; SKOUMALOVÁ, Hana. Surface and Deep Cases. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING '92), Nantes, 23-28 August 1992, Vol. 2*. Nantes: Univesity of Nantes, 1992. Pp. 885-888.
- PANEVOVÁ, Jarmila. (1996). More Remarks on Control. In Prague Linguistic Circle Papers, Vol.2. Ed. Hajičová, Eva; Leška, Oldřich, Sgall, Petr; Skoumalová, Zdena. Amsterdam; Philadelphia: J.Benjamins Publ. House, 1996. Pp. 101-120.
- PANEVOVÁ, Jarmila. Ellipsis and zero elements in the structure of the sentence. In: *Tipologija, grammatika, semantika. K 65-letiju Viktora Samuiloviča Chrakovskogo*. Sankt-Peterburg: Nauka, 1998. S. 67-76.
- PANEVOVÁ, Jarmila; HAJIČOVÁ, Eva; SGALL, Petr (2001): Manuál pro tektogramatické značkování (III. verze, prosinec 2001), <http://ufal.mff.cuni.cz/publications/year2001/MN+dodat.doc>.
- PANEVOVÁ, Jarmila; HAJIČOVÁ, Eva; SGALL, Petr: Tectogramatics in corpus tagging. In *Perspectives on Semantics, Pragmatics, and Discourse; A Festschrift for Ferenc Kiefer. Pragmatics and Beyond new Series, Vol. 90*. Eds. Kenesei, I.; Harnish, R. M. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2001. ISBN 90-272-5109-6. Pp. 294-299.
- PANEVOVÁ, Jarmila; HAJIČOVÁ, Eva; SGALL, Petr: K nové úrovni bohemistické práce: Využití anotovaného korpusu. *Slovo a slovesnost*, roč. 63, č. 3, 4, 2002. S. 161-177; 241-262.
- PETKEVIČ, Vladimír. A new formal specification of underlying representations. *Theoretical Linguistics* vol. 21, 1995. Pp. 7-61.
- SGALL, Petr. Underlying structure of sentences and its relations to semantics. In *Wiener Slawistischer Almanach, Sonderband 33. Festschrift für Viktor Julevič Rozencvejš*. Ed. by Tilmann Reuther. Wien: 1992. Pp. 273-282.
- SGALL, Petr; HAJIČOVÁ, Eva; BURÁŇOVÁ, Eva. *Aktuální členění věty v češtině*. Praha: Academia, 1980.
- SGALL, Petr; HAJIČOVÁ, Eva; PANEVOVÁ, Jarmila. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel; Praha: Academia, 1986.
- SGALL, Petr. Remarks on the Semantics of the Focus. In *Recent Topics in Mathematical and Computational Linguistics (papers in honour of S. Marcus)*. Eds. Martín-Vide, C.; Paun, G. Bucuresti: Editura Academiei Romane, 2000. ISBN 973-27-0770-4. Pp. 271-278.

- SGALL, Petr. English Syntax in Functional Generative Description, Topic-focus articulation (information structure) of the sentence, Syntax and semantics. In *Rudiments of English Linguistics*. Ed. Štekauer, Pavol. Prešov: Slovacontact, 2000. ISBN 80-88876-04-4. Pp. 225-265.
- SGALL, Petr (2001): Functional Generative Description, Word Order and Focus. *Theoretical Linguistics*, no. 27, 2001. Pp.3-19.
- SGALL, Petr (2002): Underlying Structures in Annotating Czech National Corpus. In *Current issues in formal Slavic linguistics*. Eds. Zybatow, G.; Junghanns, U.; Mehlhorn; G.; Szucsich L. Frankfurt/Main: Peter Lang, 2001. Pp. 499-505.
- SGALL, Petr: Topic-Focus Articulation in Corpus Annotation. In *Natural language processing between linguistic inquiry and system engineering*. Eds. Menzel, Wolfgang; Vertan, Christina. Iasi: Editura Universitatii "Alexandru Ioan Cuza", 2003. Pp. 95-101.
- SGALL, Petr: Dynamics in the meaning of the sentence and of discourse. In *Meaning: The Dynamic Turn*, Oxford: Elsevier Science Ltd., 2003. Pp. 169-184.
- ŠMILAUER, Vladimír. *Novočeská skladba*. Praha: Mikota, 1947.
- VESELÁ, Kateřina; PETEREK, Nino; HAJIČOVÁ, Eva. Topic-Focus Articulation in PDT: Prosodic Characteristics of Contrastive Topic. *Prague Bulletin of Mathematical Linguistics*, no. 79, 2003. Pp. 5-22.