

Redukční analýza s operacemi delete a shift a analytická rovina Pražského závislostního korpusu (PDT).

M. Plátek, K. Oliva

Rozšíření a zpřesnění prezentace z ITATu 2015. Pracovní
verze.

Úvod: tři typy redukční analýzy (RA)

Tato prezentace je neformální, se snahou o přesnost. Má ukázat souvislosti mezi metodou redukční analýzy a analytickou rovinnou Pražského závislostního korpusu (PDT). Odlišíme (nově proti sborníkovému příspěvku ITATu) **úplnou povrchovou redukční analýzu (UPRA)** od **redukční analýzy regulované A-stromy (PRAD)**, tj. stromy (skoro) ve tvaru **analytických stromů Pražského závislostního korpusu (PDT)**. UPRA, PRAD používají jako elementární operace jen delete a shift. Pomocí PRAD zavedeme **indukovanou redukční analýzu A-stromů (IPRAD)**. IPRAD používá jako základní operace také jen delete a shift. Delety skládá pouze do dvou typů složených operací: "odtržení podstromu" a "vytržení podstromu", které vysvětlíme pomocí příkladů.

Nejprve se zaměříme na varianty redukční analýzy vět jen se **závislostmi (podřízenostmi)**, upřesníme základní pojmy, pak rozšíříme pozorování o **vícenásobné koordinace a vložené koordinace**.

Úvod: tři typy redukční analýzy (RA)

Upozorníme na **typické žádoucí vlastnosti** redukční analýzy čistě závislostních vět i na **postupnou relaxaci** těchto vlastností při pozorování obou typů koordinací. Budeme pozorovat dodržování těchto vlastností na studovaném materiálu.

Jednou takovou žádoucí vlastností je, že stejně jako pro závislosti, tak i pro koordinační jevy, **velikost redukcí a neredukovatelných vět (analytických stromů)** při UPRA, PRAD i IPRAD zůstává **neveliká** a hlavně **nezávislá na velikosti vět a analytických stromů (A-stromů)**. To je vlastnost, která umožňuje přímé modelování (simulaci) redukční analýzy konečnými formálními prostředky (např. restartovacími automaty), které používají operace delete a shift.

Podrobnější důvody pro naše snažení

Prezentujeme postup, kterým lze získat podklady pro automatizaci tří variant povrchové redukční analýzy, které jsou založené na operacích delete a shift. Očekáváme, že tyto podklady budeme schopni adekvátně převést na instrukce vhodných typů restartovacích automatů a tuto adekvátnost budeme schopni prokázat pomocí zde zavedených pojmů.

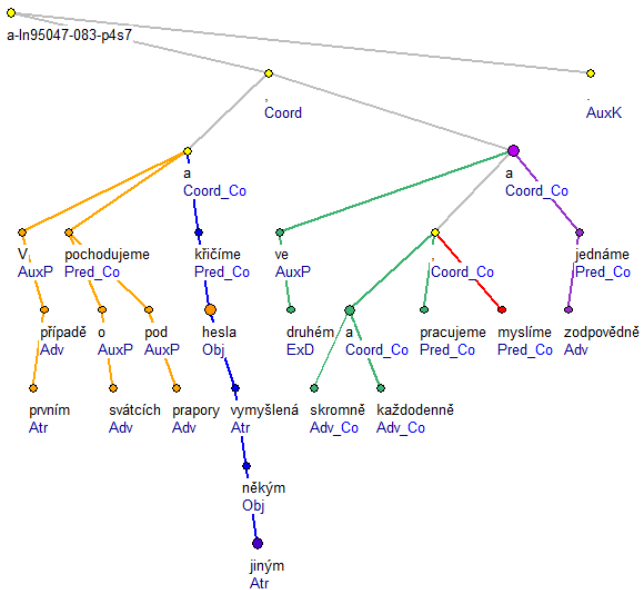
Pomocí variant redukční analýzy máme tak možnost získat verifikované podklady pro počítačovou (formální) gramatiku povrchové syntaxe češtiny, podklady pro syntaktické analyzátoři češtiny a důležitou část podkladů pro automatické vyhledávání syntaktických chyb a jejich nápravu.

Domníváme se, že pomocí automatizace redukční analýzy lze získat užitečné podklady pro učení (statistických) parserů.

Aplikacemi variant redukční analýzy lze kontrolovat konzistenci značkování v PDT a správnost (značkování) jednotlivých A-stromů z PDT. Při přípravě tohoto příspěvku jsme nějaké chyby (nekonzistence) v autentických A-stromech našli.

Náš postup je v tomto příspěvku aplikován na češtinu a analytickou rovinu PDT. Je však dostatečně obecný na to, aby mohl být aplikován na jiné jazyky a jejich tree-banky.

Autentický A-strom bez morfologických značek.



U A-stromu je podstatná **stromová struktura** s **orientovanými hranami** a **horizontální uspořádání uzlů**.

Orientace: cesty v A-stromě vedou **od listů ke kořeni**.

Uzly reprezentují **syntakticky značované** lexikální a interpunkční jednotky.

Stromová struktura zachycuje syntaktické vztahy (v následujícím jen **závislosti a koordinace**).

Horizontální uspořádání uzlů vyjadřuje slovosled.

A-stromy nesou úplnou informaci o větách, nad kterými jsou sestrojeny i o jejich označování.

O větě se značováním, která je popsána A-stromem T , říkáme, že je **projekcí A-stromu T** .

A-strom reprezentuje danou korektní větu tak, jak ji anotátor pochopil.

Anotátor používá následující dovednosti :

(1) Umí zkontrolovat korektnost věty a korektně označkovat její slova (i interpunkční znaménka) syntaktickými a morfológickými značkami. Umí zkontrolovat korektnost daného značkování.

(2) Z takto označované věty umí sestrojít korektní A-strom. Umí zkontrolovat korektnost daného A-stromu.

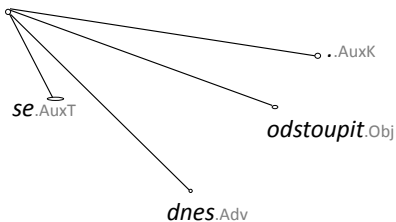
Na dalších slídách budeme ukazovat, jak na **označkových větách** provádět **úplnou redukční analýzu (UPRA)**, a jak pro dané **A-stromy** udělat **regulovanou redukční analýzu (PRAD)** i **indukovanou redukční analýzu A-stromu (IPRAD)**.

Předpokládáme pro to (idealizované) anotátorské dovednosti.

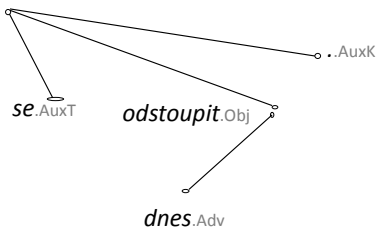
Označovaná věta (0) a její A-stromy $T0_a$ a $T0_b$

(0) *Rozhodl.Pred se.AuxT dnes.Adv odstoupit.Sb ..AuxK*

Rozhodl.Pred



Rozhodl.Pred



Zásady základních variant redukční analýzy.

Zásady pro UPRA (úplnou povrchovou redukční analýzu) a **zásady pro PRAD (redukční analýzu řízenou A-stromem):**

- (i) UPRA i **PRAD** jsou složeny z jednotlivých redukcí; redukce používají dva typy operací : (1) vypuštění (delete) a (2) shift (přesun). To znamená, že tvary jednotlivých slov (i interpunkčních znamének), jejich morfologické charakteristiky i jejich syntaktické kategorie se nemění během jednotlivých redukcí;
- (ii) gramaticky správná věta (včetně značkování i morfologických charakteristik) musí zůstat správná i po redukci;
- (iii) uvažujeme jen nezmenšitelné redukce, t.j. vynecháme-li z libovolné redukce jednu, či více operací, nastane porušení principu zachování správnosti (ii);

Zbývající zásady obou větných redukčních analýz.

- (iv) předložkové vazby (např. 'o otce'), se v redukci vynechávají celé (jinak je možný posun významu, často i změny v pádech);
- (v) redukce používají operaci shift jenom v případech vynucených principem zachování korektnosti, tedy v případech, kdy vynechání shiftu by vedlo k nekorektnímu větnému slovosledu;
- (vi) A-strom věty po redukci zachovává stromovou strukturu A-stromu věty před redukcí. Přesněji: **zachovává kořen a změny ve stromové struktuře jsou způsobeny odstraněním souvislých začátků cest (začíná se od listů).**
- (vii) úplná povrchová redukční analýza (UPRA) obsahuje všechny možné redukce, splňující zásady (o) až (v);
- (viii) **redukční analýza regulovaná A-stromem (PRAD) obsahuje všechny možné redukce, splňující zásady (o) až (vi).**

Indukovaná redukční analýza A-stromu (IPRAD)

IPRAD zavádíme následujícím způsobem:

Mějme korektní A-strom T . IPRAD A-stromu T pracuje s množinou A-stromů, které byly použity (vytvořeny) během PRAD podle A-stromu T . Jako základní operace používá operace delete a shift. Připomeňme si, že po operaci delete je v obecném případě nutné zajistit (původní) souvislost výsledného stromu pomocí přidání jistých hran. Počet přidanych hran určuje složitost provedení operace delete. Zavedením hierarchie si delete přestaly být rovné.

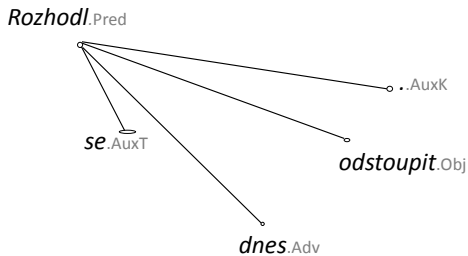
Mezi redukcemi IPRAD A-stromu T a redukcemi PRAD podle T požadujeme vzájemně jednoznačné zobrazení : T_1 je v IPRAD redukováno na T_2 právě tehdy, když projekce A-stromu T_1 je v PRAD redukována na projekci A-stromu T_2 .

Ukážeme na následujících příkladech A-stromů, že na základě předchozí definice IPRAD používá redukce založené na posloupnostech základních operací delete a shift pro A-stromy, přičemž IPRAD z operací delete skládá právě dva typy složených operací. Prvnímu typu říkáme **odtržení podstromu**, druhému říkáme **vytržení podstromu**. Uvidíme, že operace typu vytržení podstromu jsou vzácnější a jejich použití je zde vynuceno pouze redukcí vložené koordinace.

Pro popis složených operací můžeme využít složitosti operace delete. Ve většině případů IPRAD využívá deletey se složitostí 0 (u odtržení stromu), jen u vytržení stromu poslední delete má složitost 1 (podle naší metodiky na našich příkladech).

Věta (0) a A-strom $T0_a$ jako vstupy do UPRa a PRAD

(0) *Rozhodl.Pred* *se.AuxT* *dnese.Adv* *odstoupit.Sb* *..AuxK*



Říkáme, že věta (0) je projekcí A-stromu $T0_a$. Je i projekcí A-stromu $T0_b$.

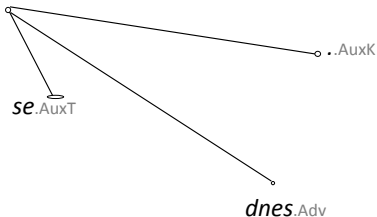
Větu (0) lze bez regulace i podle $T0_a$ přímo zredukovat dvěma způsoby, a to buď na větu (0a1), nebo na větu (0b1), viz pokračování.

Redukce věty (0a1) na (0a2) a A-stromu T_{0a1} na T_{0a2}

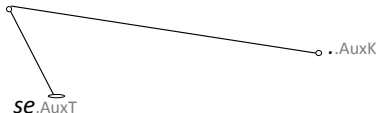
(0a1) *Rozhodl.Pred se.AuxT dnes.Adv ..AuxK*

(0a2) *Rozhodl.Pred se.AuxT ..AuxK*

Rozhodl.Pred



Rozhodl.Pred

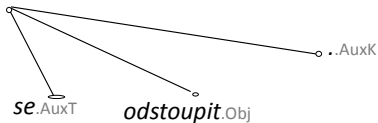


Redukce věty (0b1) na (0a2) a A-stromu T_{0b1} na T_{0a2}

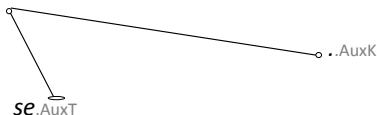
(0b1) *Rozhodl.Pred se.AuxT odstoupit.Sb ..AuxK*

(0a2) *Rozhodl.Pred se.AuxT ..AuxK*

Rozhodl.Pred

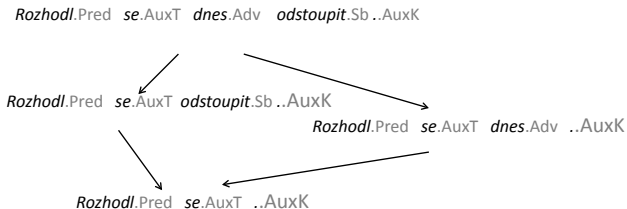


Rozhodl.Pred



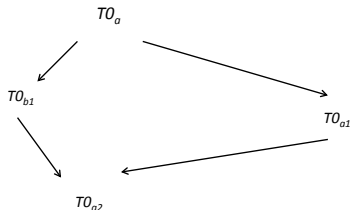
UPRA věty (0) a PRAD podle $T0_a$

Z předchozích slíd vyplývá, že UPRA věty (0) a PRAD podle $T0_a$ jsou stejné.



Všechny možné (nezmenšitelné) redukce věty (0).

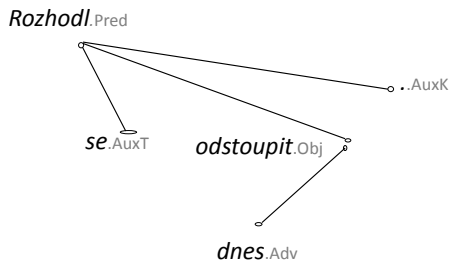
Schema indukované redukční analýzy (IPRAD) A-stromu $T0_a$



IPRAD je dáno A-stromy a jejich redukcemi, které byly použity pro výběr redukcí během PRAD. Schema IPRAD má stejný tvar jako schema PRAD.

Věta (0) a A-strom $T0_b$ jako vstup do PRAD

(0) *Rozhodl.Pred se.AuxT dnes.Adv odstoupit.Sb ..AuxK*



Větu (0) lze podle $T0_b$ přímo zredukovat jediným způsobem, a to na větu (0b1).

PRAD a IPRAD podle $T0_b$

PRAD podle $T0_b$ se liší od UPRA věty (0). $T0_b$ nese více informace než označkování věty (0). Rozdíl je v tom, že slovo **dnes** je v PRAD podle $T0_b$ povinně vypuštěno jako první.

Rozhodl.Pred se.AuxT dnes.Adv odstoupit.Sb ..AuxK



Rozhodl.Pred se.AuxT odstoupit.Sb ..AuxK



Rozhodl.Pred se.AuxT ..AuxK

$T0_b$



$T0_{b1}$



$T0_{a2}$

Označkováná věta (1) a A-strom T_{1_1} nad větou (1)

(1) *Vidí.Pred otce.Obj matky.Atr souseda.Atr ..AuxK*

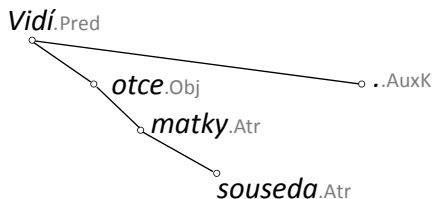
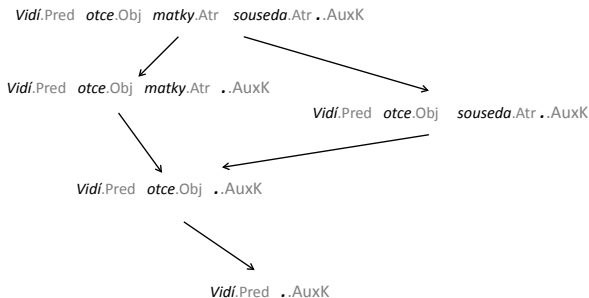


Figure: Čistě závislostní A-strom T_{1_1} nad větou (1).

UPRA věty (1)

Tento příklad ukazuje, že A-strom $T1_1$ nese více (významové) informace než pouhé označkování věty (1). A to přes to, že $T1_1$ je k danému označkování věty (1) jediným stromem.



Schema všech možných (nezmenšitelných) redukcí věty (1).

Redukce A-stromu $T1_1$ na $T1_2$ (odtržením podstromu)

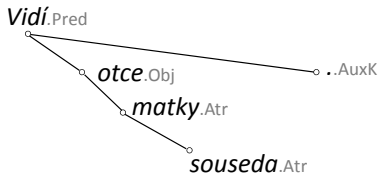


Figure: A-strom $T1_1$



Figure: A-strom $T1_2$

Redukce A-stromu $T1_2$ na $T1_3$ (odtržením podstromu)



Figure: A-strom $T1_2$

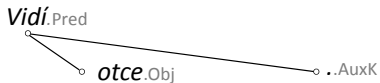


Figure: A-strom $T1_3$

Redukce A-stromu $T1_3$ na $T1_4$ (odtržením podstromu)

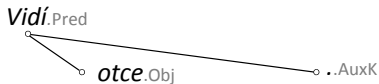
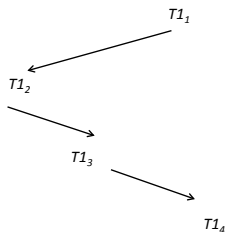
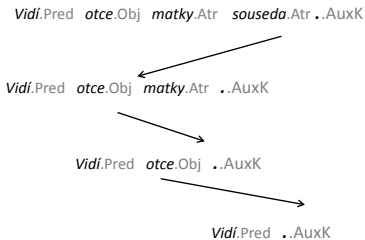


Figure: A-strom $T1_3$



Figure: A-strom $T1_4$

PRAD a IPRAD podle A-stromu $T1_1$.



Věta se závislostmi (2) a A-strom $T2_1$ nad větou (2)

(2) Petr.Sb se.AuxT bojí.Pred o.AuxP otce.Obj ..AuxK

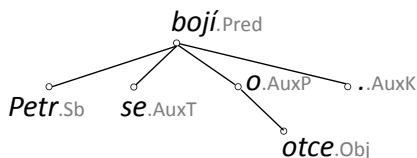


Figure: Čistě závislostní A-strom $T2_1$ nad větou (2).

Tento příklad představuje redukce se shiftem. Ilustruje také tři typy syntaktických chyb.

UPRA věty (2)

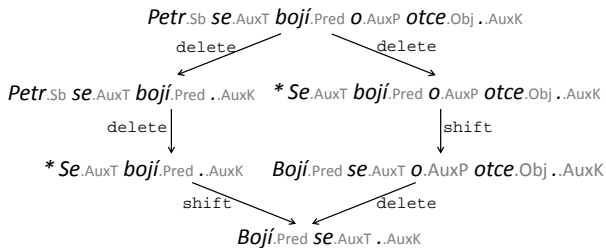


Figure: Schema UPRA pro větu (2).

Lze ukázat, že tato UPRA je rovna PRAD podle A-stromu $T2_1$. Redukce se shiftem jsou ve schématu rozloženy do dvou kroků, aby ilustrovaly vznik a nápravu chyb ve slovosledu. Pro zkrácení ukážeme jen redukce A-stromů podle levé větve schématu.

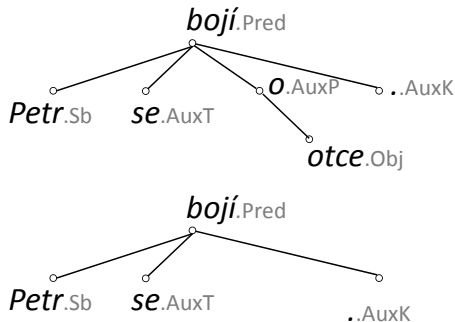


Figure: $T2_1$ je redukováno odtržením podstromu na $T2_2$.

Vynecháním "slova" *otce* získáme větu se syntaktickou chybou.
Vynecháním "o" získáme nekonzistenci v morfologickém značkování.

Redukce $T2_2$ na $T2_4$.

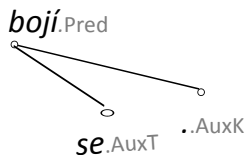
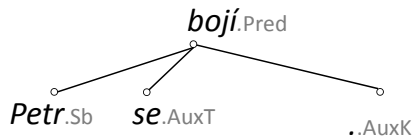


Figure: $T2_2$ je redukován na $T2_4$ pomocí odtržení podstromu a shiftu

A-strom $T2_4$ je neredukovatelný, vynecháním kteréhokoliv slova (uzlu) bychom získali větu se syntaktickou chybou, napravitelnou doplněním na původní větu.

Věta (3) a A-strom $T3_1$ (s koordinacemi)

(3) *Je.Pred* *dědou.Obj.Co* , *.AuxX* *otcem.Obj.Co* *a.Cr*
strýcem.Obj.Co *..AuxK*

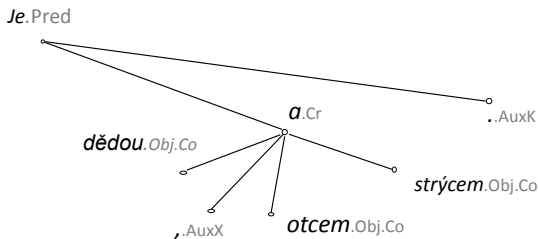


Figure: A-strom $T3_1$.

Používáme značku **Cr** místo **Coord**.

Tento příklad ukazuje redukce vícenásobných koordinací a další typ syntaktických chyb.

UPRA věty (3), věty s vícenásobnou koordinací

Je.Pred dědou.Obj.Co ,.AuxX otcem.Obj.Co a. Cr strýcem.Obj.Co ..AuxK

Je.Pred otcem.Obj.Co a.Cr strýcem.Obj.Co ..AuxK

Je.Pred dědou.Obj.Co a.Cr strýcem.Obj.Co ..AuxK

Je.Pred dědou.Obj.Co a.Cr otcem.Obj.Co ..AuxK

Je.Pred ..AuxK

shift

UPRA pro větu s trojnásobnou koordinací.

Ukážeme, že UPRA věty (3) je také PRAD podle $T3_1$.

Redukce vícenásobné koordinace v A-stromech.

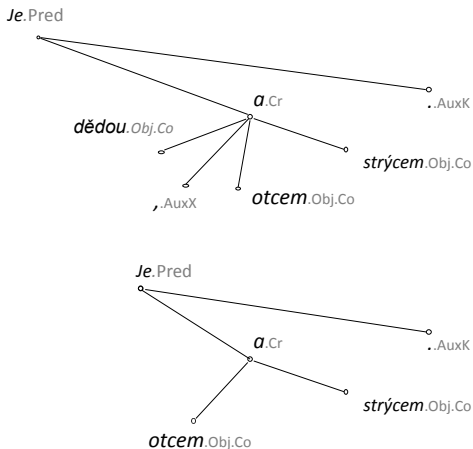


Figure: T_3 redukováno na T_3_2 pomocí dvou odtržení podstromu. Jedním odtržením podstromu bychom získali větu (A-strom) se syntaktickou chybou.

Redukce $T3_1$ na $T3_3$.

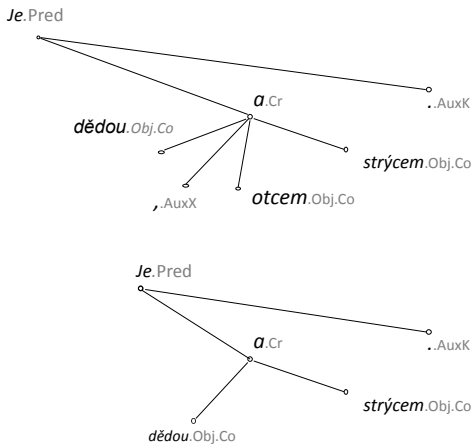


Figure: $T3_1$ redukováno na $T3_3$. Redukce se podobá předchozí redukci.

Redukce $T3_1$ na $T3_4$.

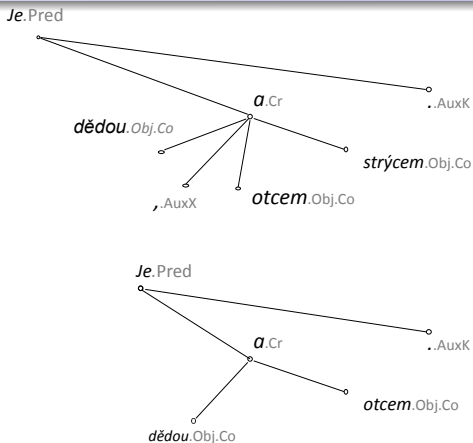


Figure: $T3_1$ redukováno na $T3_4$ pomocí dvou odtržení podstromu a shiftem.

Vynecháním jedné ze tří operací bychom získali větu se syntaktickou chybou.

Redukce (vypuštění) závislé koordinace.

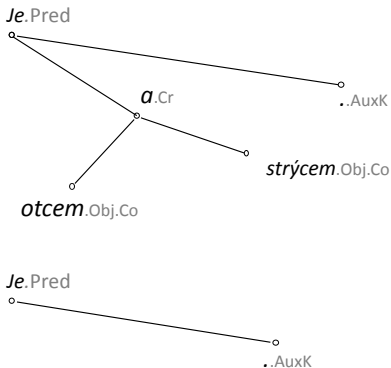


Figure: T_{3_2} redukováno na T_{3_5} jedním odtržením podstromu.

Podobně jedním odtržením podstromu lze na T_{3_5} redukovat T_{3_3} a T_{3_4} . Zmenšením redukce bychom získali syntaktickou chybu, nebo nekonzistenci ve značkování.

Věta (4) a A-strom T_{4_1} s koordinací závislou na řídicí koordinaci.

(4) *Skromně*.Adv.Co *a*.Cr *denně*.Adv.Co *pracujeme*. Pred.Co
a.Cr *jednáme*.Pred.Co ..AuxK

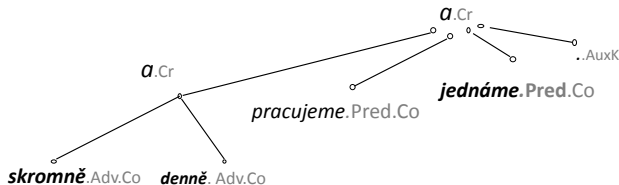


Figure: A-strom T_{4_1}

UPRA věty (4).

Skromně.Adv.Co *a*.Cr *denně*.Adv.Co *pracujeme*. Pred.Co *a*.Cr *jednáme*. Pred.Co ..AuxK



pracujeme. Pred.Co *a*.Cr *jednáme*. Pred.Co ..AuxK

Figure: UPRA závislé koordinace na řídící koordinaci.

Tato UPRA je rovna PRAD podle A-stromu $T4_1$.

Znázorňuje jedinou redukci, která odstraňuje koordinovaná příslovečná určení, která rozvíjejí koordinované predikáty. Zmenšením redukce bychom získali syntaktickou chybu, nebo nekonzistenci ve značkování.

Redukce $T4_1$ na $T4_2$.

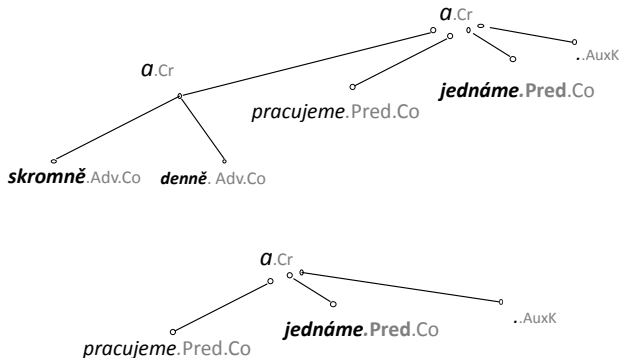


Figure: Redukce $T4_1$ na $T4_2$ pomocí jednoho odtržení podstromu.

Věta (5) a A-strom $T5_1$ s vloženou koordinací.

(5) *Pracujeme.Pred.Co a.Cr.Co myslíme .Pred.Co i.Cr
jednáme.Pred.Co ..AuxK*

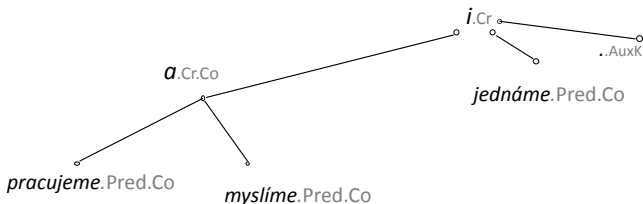


Figure: A-strom s koordinací vloženou do koordinace.

UPRA věty (5).

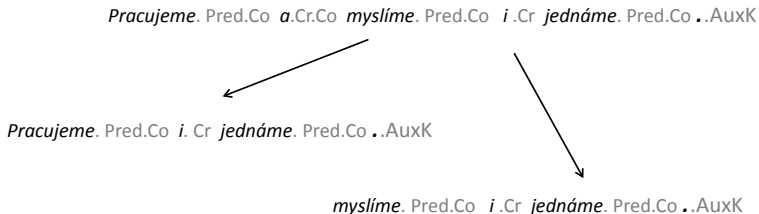


Figure: UPRA věty s vloženou koordinací.

Ukážeme, že PRAD podle $T5_1$ je rovna UPRA věty (5).
Zmenšením redukce získáme syntaktickou chybu, nebo nekonzistenci v označování.

Redukce $T5_1$ na $T5_2$ vytržením podstromu.

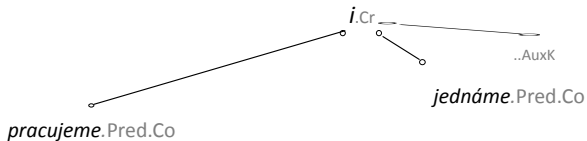
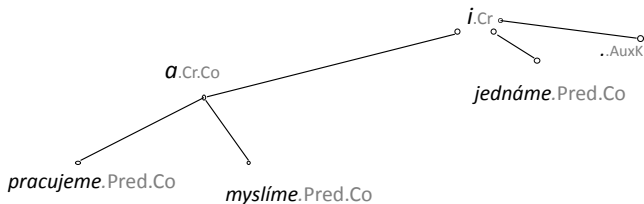


Figure: $T5_1$ redukováno na $T5_2$.

Redukce $T5_1$ na $T5_3$ vytržením podstromu.

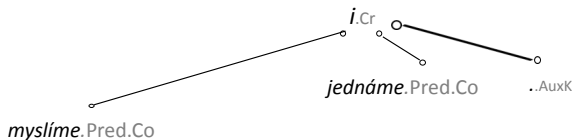
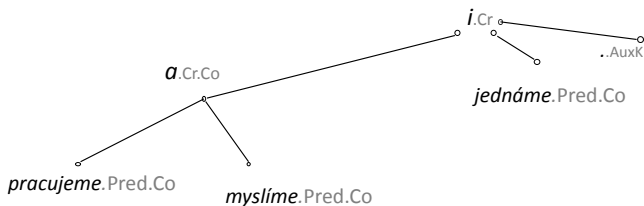


Figure: $T5_1$ redukováno na $T5_3$.

Úvod k souhrnu výsledků a pozorování

Od redukční analýzy přirozených jazyků očekáváme plnění jistých význačných vlastností, které najdeme na následujících slídách. Toto plnění zkoumáme a omezujeme se přitom na závislostní a koordinační jevy v češtině, anotované způsobem obvyklým u analytické roviny PDT. Do značné míry jsou (měly by být) tyto vlastnosti zaručeny kvalitou použitého syntaktického značkování. Ověřujeme tedy, kromě jiného, také kvalitu tohoto značkování. Naší snahou je ukázat, že se základními operacemi **delete a shift** pro všechny tři varianty RA a jen se složenými operacemi **odtržení podstromu a vytržení podstromu** pro IPRAD jsme schopni, tyto význačné vlastnosti garantovat. Z rozboru prezentovaných příkladů vyplývá, že pro splnění celého souhrnu těchto vlastností, nelze žádnou z těchto operací opomenout.

Souhrn výsledků a pozorování

Očekávaných sedm význačných vlastností:

- (1) Pro všechny tři typy redukční analýzy (UPRA, PRAD, IPRAD) je splněno, že všechny větve redukčních analýz stejné věty jsou stejně dlouhé. Nepoužijeme-li operaci shift, tuto vlastnost pro některé věty ztrácíme.
- (2) Pro UPRA (a PRAD) platí, že v redukční analýze jedné věty dvě různé redukce vypouštějí na množině disjunkčních pozic. Odtud vyplývá podobné pozorování i pro IPRAD. Pokud bychom nepoužívali značkování, tak toto pozorování nemusí být splněno.
- (3) UPRA a PRAD vět(y) se závislostmi a bez koordinací mají jednoznačně určenou neredukovatelnou větu (RA tvoří svaz) a vypouštěcí část RA v jednotlivých větvích se liší jen pořadím provedených redukčních vypuštění. Různé větve jedné RA se však mohou lišit množinou provedených přesunů. Neuvažujeme-li operaci shift, tuto vlastnost pro některé věty ztrácíme.

Souhrn výsledků a pozorování

(4) RA vět s koordinacemi nemusí používat v různých větvích stejná vypuštění a mohou mít více neredukovatelných vět než jednu.

(5) Neredukovatelné věty PRAD jsou i neredukovatelnými větami v UPRA.

(6) Horní odhad velikosti redukcí a neredukovatelných vět u všech tří variant RA je nezávislý na velikosti redukovaných vět a A-stromů a není nijak velký.

Nepoužíváme-li operaci shift, tuto vlastnost pro některé věty ztrácíme. Nepoužíváme-li operaci vytržení podstromu, ztratíme tuto vlastnost pro A-stromy s vloženou koordinací. Povolíme-li jen redukce s jediným odtržením podstromu, ztratíme tuto vlastnost pro A-stromy s vícenásobnou koordinací.

(7) Pro sestrojení A-stromu k dané větě stačí jít proti směru jedné (libovolné) větve indukované redukční analýzy. Tedy k provedení (povrchové) syntaktické analýzy stačí postupovat podle jediné větve IPRADu.

Souhrn výsledků a pozorování

Na této slídě se snažíme načrtnout, jak potvrzovat pozorování (6).

Pracujeme s (dříve ověřovaným) předpokladem, že tato skutečnost platí pro věty se závislostmi, které nemají koordinace.

Dále se domníváme, že jsme pomocí příkladů probrali všechny typy redukcí, které jsou v PDT svázané se značkami pro koordinace.

Z těchto příkladů vidíme, že typ redukce, který odpovídá finálnímu odstranění koordinace, nemůže být o mnoho větší (asi o jedna) než dvojnásobek velikosti maximální čistě závislostní redukce (jde o velikost vypouštění). Ostatní typy redukcí svázané s koordinacemi budou mít horní odhad zhruba stejný, jako mají čistě závislostní redukce.

Tyto předpoklady bychom měli podrobit diskusi (něco bude dál).



Ostatní vlastnosti:

Uvažujme vypouštěcí a přesunovací složitost (maximální počet slov vypuštěných resp. přesunutých v jedné redukci).

Vypouštěcí složitost čistě závislostních vět odhadujeme číslem 7, přesunovací složitost číslem 5.

Na základě našich pozorování by se vypouštěcí složitost vět s koordinacemi měla odhadnout zhruba dvojnásobkem + 1 vypouštěcí složitosti vět bez koordinací, tedy např. číslem 15.

Horní odhad přesunovací složitosti by koordinace neměly měnit.

Oba typy složitosti se mohou u různých jazyků, či tree-banků (nebo i anotátorů) lišit.

Vlastnosti IPRADu:

Redukce IPRADu čistých závislostí používají právě jedno odtržení podstromu a případně ještě nějaké shifty.

Redukce IPRADu, která snižuje násobnost koordinace, používá dvě odtržení podstromu a případně ještě nějaké shifty.

Redukce IPRADu, která odstraňuje dvojnásobnou koordinaci, používá jediné odtržení podstromu a případně nějaké shifty. V tomto případě na použití shiftů (zatím) nemáme příklad.

Redukce IPRADu, která odstraňuje vloženou koordinaci do koordinace, používá jedno vytržení podstromu. V tomto případě asi ani nebude existovat příklad na použití shiftů.

○ redukční analýze a struktuře A-stromu:

Mějme dvě hrany h_1, h_2 v A-stromě T . Z h_1 vede cesta do h_2 . Pak redukce IPRADu A-stromu T odstraňující h_1 nemůže být provedena později než redukce odstraňující h_2 .

Mějme dva uzly u_1, u_2 v A-stromě T neobsahujícím vložené koordinace. Z u_1 vede cesta do u_2 . Pak redukce IPRADu A-stromu T odstraňující u_1 nemůže být provedena později než redukce odstraňující u_2 .

Mějme dva uzly u_1, u_2 v A-stromě T . Z u_1 nevede cesta do u_2 a nevede ani opačná cesta. Pak redukce IPRADu A-stromu T odstraňující u_1 a u_2 mohou být provedeny v obou pořadích (pokud jsou dvě).

Mějme dvě slova s_1, s_2 ve větě V a UPRA věty V redukuje s_1 vždy dříve (v každé větvi UPRA) než s_2 . Pak z s_2 nevede cesta do s_1 v žádném A-stromě nad větou V .

O redukční analýze a struktuře A-stromu:

Pozorování na předchozí slídě vyplývají z definic IPRAD a UPRA. Upřesňují a rozšiřují tradiční intuitivní výklad souvislostí mezi redukční analýzou a mezi (čistě) závislostními stromy.

Co ještě chybí ?

Máme možnost předvedené typy redukční analýzy zjemnit pomocí speciálního přepisování, které by např. umožňovalo odstraňovat koordinační značky.

Varianty redukční analýzy je možné parametrizovat množinou použitých operací. To umožňuje zkoumat jejich vlastnosti a především vlastnosti jazyků, které obvyklé prostředky formálních jazyků zkoumat neumožňují.

Co chybí ?

Chybí analyzované příklady na složené věty s podřízenými větami a jejich kombinace s koordinacemi a diskuse k nim. Těmito příklady získáme spodní odhady velikosti redukcí.

Např. Věra, jejíž matka se začala bát a jejíhož otce se nepodařilo vyléčit, nezpanikařila.

Co ještě chybí ?

Zmínit, že UPRA pro danou větu dává v jistém smyslu úplný systém povrchových složek a tyto složky mohou být nespojité.

Důležitý parametr (složitosti věty) : Počet redukcí ve větvi redukční analýzy.

Máme praktický odhad tohoto parametru (nejdelší větou v tree-banku). To dává konečný jazyk.

Nemáme teoretický horní odhad tohoto parametru : To dává nekonečný jazyk.

Redukční analýza dává možnost definovat různé míry stejně relevantní pro konečné i nekonečné jazyky.

Kromě vypouštěcí a přesunovací složitosti např. míry nespojitosti redukcí, neprojektivity redukcí a pod.

Tady končí pracovní verze naší prezentace